

Query Match	Best Local Match	Similarity	Score	3107.8 ;	DB 4 ;	Length 8698 ;
Matches 3790 ;	Conservative	0 ;	Mismatches 882 ;	Indels 53 ;	Gaps 1	
QY	1	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAAAGGTCCGC	60			
Db	1	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCCGCGGCGCAAAAGGTGCGC	60			
QY	61	AGACGGCAGAGCTCTGCTCTGCTCCGCGCCCAACGAGCGGAGCGCGGAGAGAGGAGTG	120			
Db	61	CGACGCGCCGCGGCTTTGCCCCGCGCGCTCTCAGTGACGACGACGCGCGGAGAGAGGAGATG	120			
QY	121	GGCAATCCATCACTAGAGGGGTATATGCGAAGCGCTCCACAGCTGCGCGCTCAGCGCTGA	180			
Db	121	GCCAACTTCATCACTAGAGGGT-----TCTCGAGAGGGGTGAGTCTGTGA	166			
QY	181	CGTAATTAACGTATAGAGG---GAGTGTGCTGTATTAGCTGTCACTGAGTGTCTTTGCG	23			
Db	164	CGTGAATTACGTATAGAGGTCGTGTATTAGAGGTACAGTGAAGTG-TTTTGCG	222			
QY	238	GACATTTTGGACACCACTGCGCATTTAGGATATATATGCGCGAGTGAAGGACAGAT	29			
Db	223	GACATTTTGGACACCATATGCTGTACAGCTGGGATATTATTAAGCCCGAGTGAAGACACGAGGGT	287			
QY	298	CTCCATTTTGG-AACCGAAATTGAACAGACGACGACCATGCGGGGCTTTACAGAGATCG	35			
Db	283	CTTCATTTTGAAGCGGAGGTTTGAACGCGACGCGCATGCGGGGTTTACAGAGATG	344			

QY 357 TGATCAAGGTCGAGGACCTGAGCAGACCTGCGGGCATTTCTGACTGTTTGA 416
DB 343 TGATTAAGTCCCAAGGACCTTGAAGCATCTGCGGACATTTCTGACAGCTTTTGA 402
QY 417 GCTGGTGGCCGAGAGAAATGAGAGTGGCCCCCGGATTTGACATGATCTGATCTGA 476
DB 403 ACTGGGTGGCCGAGAGAAATGAGAGTGGCCCCCGGATTTGACATGATCTGATCTGA 462
QY 477 TTGAGCAGGACCCCTGACCTGGCCGAGAACTGACAGCGGAACTTCTGTGCAATGGC 536
DB 463 TTGAGCAGGACCCCTGACCTGGCCGAGAACTGACAGCGGAACTTCTGTGCAATGGC 522
QY 537 GCGCGTGAAGTGAAGCCCGGAGGCGCTCTTTGTTGATGAGAGAGGAGAGTCT 596
DB 523 GCGGTGAGTGAAGGCGCGGAGGCGCTTTTCTTTGTCATTTGAGAGAGAGAGTCT 582
QY 597 ACTTCACTCTCATATTTCTGGTGAAGCAGCGGGGTCATATTCATGCTGAGGCGCT 656
DB 583 ACTTCACTCACTGCTGCTGAGAAACCAAGGGGTCATATTCATGCTGAGGCGCT 642
QY 657 TCTGATGATGATGAGGACAGCTGTCAGACCATCTACCGGGGATGAGGCGGACCC 716
DB 643 TCTGATGATGATGAGGACAGCTGTCAGACCATCTACCGGGGATGAGGCGGACCC 702
QY 717 TGCCCACTGGTGGCGGTGACCAAGACGCTAATGCGCGCGGAGGAGCAAGGTGG 776
DB 703 TGCCCACTGGTGGCGGTGACCAAGACGCTAATGCGCGCGGAGGAGCAAGGTGG 762
QY 777 TGAGCAGTGTATCATCCCACTAATCTCTGCGGACAGCTGAGCGGAGCTGAGTGG 836
DB 763 TGAGTGTGTATCATCCCACTAATCTCTGCGGACAGCTGAGCGGAGCTGAGTGG 822
QY 837 CGTGAATCAATGAGAGATATTAAGCGCTGTTGAACTTGGCGGAGGAGCAAGCGC 896
DB 823 CGTGAATCAATGAGAGATATTAAGCGCTGTTGAACTTGGCGGAGGAGCAAGCGC 882
QY 897 TCGTGGCGCAGCAGCTGACCCAGCTGACGAGCCAGAGCAGAGCAAGAGAGATCTGA 956
DB 883 TCGTGGCGCAGCAGCTGACCCAGCTGACGAGCCAGAGCAGAGCAAGAGAGATCTGA 942
QY 957 ACCCAATTTCTGACGCGCTGTCTATCGGGTCAAAAACCTCGCGGCTACATGAGTGG 1016
DB 943 ATCCCAATTTCTGACGCGCTGTCTATCGGGTCAAAAACCTCGCGGCTACATGAGTGG 1002
QY 1017 TCGGGTGGCTGAGAGCCGGGGACCTGACCTCGAGAGAGAGTGAATCAAGAGAGACAG 1076
DB 1003 TCGGGTGGCTGAGAGCCGGGGACCTGACCTCGAGAGAGAGTGAATCAAGAGAGACAG 1062
QY 1077 CCTGATCATCTCCTTCAACGCGCTTCAACCTGCGGCTCCAGATCAAGGCGCTGAG 1136
DB 1063 CCTGATCATCTCCTTCAACGCGCTTCAACCTGCGGCTCCAGATCAAGGCGCTGAG 1122
QY 1137 ACAATGCGGCAAGATCATGAGCGCTGACCAATCGCGGCTGACTGCTGAGGCGCG 1196
DB 1123 ACAATGCGGCAAGATCATGAGCGCTGACCAATCGCGGCTGACTGCTGAGGCGCG 1182
QY 1197 CTCGCGCGGCAAGATCAAAAACGAGCATCTACCGGATCTGAGAGTGAAGGCTAG 1256
DB 1183 AGCGGTGAGAGCAATTTCCAGCAATGAGATTTTAAATTTTGAATCAAGGAGTACG 1242
QY 1257 AACCTGCTACGCGGCTCGCTCTTCTCGGCTGGGCGCAGAAAAGTTGCGGAGAGCG 1316
DB 1243 ATCCCAATTAAGCGGCTTCGCTCTTCTGAGATGGGCGAGAAAAGTTGCGGAGAGG 1302
QY 1317 ACACATCTGCTGTTGGGCTGACCACTACCGGAGAGACCAATGAGGAGGCGCTAG 1362
DB 1303 ACACATCTGCTGTTGGGCTGACCACTACCGGAGAGACCAATGAGGAGGCGCTAG 1352
QY 1377 CCGACGCGGCGCTTCTGAGGCTGCGTCACTGAGCAATGAGAACTTCTCTTCAATG 1436
DB 1363 CCGACGCGGCGCTTCTGAGGCTGCGTCACTGAGCAATGAGAACTTCTCTTCAATG 1422
QY 1437 ATTGCTGACAGAGTGTGATCTGTGGGAGAGGCGCAAGATGACGCGCAAGGTCTG 1496

DB 1423 ACTGTGTCAGCAAGATGTGATCTGTGTGGAGAGGAGGAGAAAGATGACCGCAAGTGTG 1482
QY 1497 AGTCCGCAAGGCAATCTTCTGCGGCGAGCAGAGTGGCGCTGGAACCAAAATGCAATGCT 1556
DB 1483 AGTCCGCAAGGCAATCTTCTGCGGCGAGCAGAGTGGCGCTGGAACCAAAATGCAATGCT 1542
QY 1557 CCGCCAGATGACCCCAACCCCGGTGATGCTGACCTTCAACCAACCAATGAGGCGCTGA 1616
DB 1543 CCGCCAGATGACCCCAACCCCGGTGATGCTGACCTTCAACCAACCAATGAGGCGCTGA 1602
QY 1617 TTGACGGAACAGACCAACCTTGAAGCAGCAGAGCGGTTGAGAGCCGATGTTCAAT 1676
DB 1603 TTGACGGAACAGACCAACCTTGAAGCAGCAGAGCGGTTGAGAGCCGATGTTCAAT 1662
QY 1677 TTGAATCAACCCCGCTTGAAGATGATCTTGGCAAGGTGACAAAGCAGAGATCAAG 1736
DB 1663 TTGAATCAACCCCGCTTGAAGATGATCTTGGCAAGGTGACAAAGCAGAGATCAAG 1722
QY 1737 AGTTCTTCCGCTGGGCGAGATCACTGACCGAGGTGGCGATGATGATGATGATGAT 1796
DB 1723 AGTTCTTCCGCTGGGCGAGATCACTGACCGAGGTGGCGATGATGATGATGATGATGAT 1782
QY 1797 AGGTTGAGGCAAAAGAACCCCGCGGATGACCGGATTAAGGAGGAGCCCAAGCGG 1856
DB 1783 AGGTTGAGGCAAAAGAACCCCGCGGATGACCGGATTAAGGAGGAGCCCAAGCGG 1842
QY 1857 CCGCCCGTGAAGTGGGATTCATGACCTGACAGCGGAGAGAGTCTCGGTGAGCTTGG 1916
DB 1843 TGGCGAGTCAAGTTGGGAGCGCATGACGTCAGACGCGGAG--AGCTTCATCACTACG 1899
QY 1917 CCGCAGGATCAAAACCAATGTTCTGTCAGCAGCGGATGCTTCAATGATGCTGTTCCCT 1976
DB 1900 CAGACAGATCAAAACCAATGTTCTGTCAGCAGCGGATGCTTCAATGATGCTGTTCCCT 1959
QY 1977 GCAAGCAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2036
DB 1960 GCAAGCAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2019
QY 2037 ACTGTTCAAGTGTCTTCCCGGGGTGAGAACTCAACCGGT---GTCAGAAAGGGA 2093
DB 2020 ACTGTTCAAGTGTCTTCCCGGGGTGAGAACTCAACCGGT---GTCAGAAAGGGA 2076
QY 2094 CGTATCGAAGATCTGTCGATCATCATCTGTCGAGGCGGAGCTCCGAGATGCTTGTCT 2153
DB 2077 CGTATCGAAGATCTGTCGATCATCATCTGTCGAGGAGGAGGAGGAGGAGGAGGAG 2133
QY 2154 CCGCTGCGATCTGTCACAGTGAACCTGATGATGATGATGATGATGATGATGATGATG 2213
DB 2134 CCGCTGCGATCTGTCACAGTGAACCTGATGATGATGATGATGATGATGATGATGATG 2193
QY 2214 AAACCAAGTATGAGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 2273
DB 2194 AAACCAAGTATGAGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 2253
QY 2274 GGCATTCGAGTGTGAGGAGCTTGAACCTGAGAGCCCGAGAGCCCAAGGCAACAGCA 2333
DB 2254 GGCATTCGAGTGTGAGGAGCTTGAACCTGAGAGCCCGAGAGCCCAAGGCAACAGCA 2313
QY 2334 AAGCAGGAGCAGCGCGGGGTCTGAGCTTCTGAGCTCAAGTACCTGAGACCTTCAAC 2393
DB 2314 AAGCAGGAGCAGCGCGGGGTCTGAGCTTCTGAGCTCAAGTACCTGAGACCTTCAAC 2373
QY 2394 GAGCTGACAGAGGAGGAGCCCGTCAACGCGGCGAGAGCGGCGGCTCGAGAGCAGCA 2453
DB 2374 GAGCTGACAGAGGAGGAGCCCGTCAACGCGGCGAGAGCGGCGGCTCGAGAGCAGCA 2433
QY 2454 GCTTACAGCAGAGGCTCAAGAGGAGTGAACCTGATCTGAGGATTAACAGAGCGGAG 2513
DB 2434 GCTTACAGCAGAGGCTCAAGAGGAGTGAACCTGATCTGAGGATTAACAGAGCGGAG 2493
QY 2514 GCGGATTTCAAGAGCGTCTGCAAGAAATGATCTTTTGGGAGCACTTGGGCGAGCA 2573

Db 2494 GCGGAGTTTCAGAGGCGCTTAAAGAAATACGTCCTTTTGGGGGCAACCTCGAGCAGCA 2553
QY 2574 GTCTTCAGAGCCAGAGGGGGTTCGAACTCTGGTCTGGTGAAGAGGCGCTAAG 2633
Db 2554 GTCTTCAGAGCCAGAGGGGGTTCGAACTCTGGTCTGGTGAAGAGGCGCTAAG 2613
QY 2634 ACAGGCTCTGGAAGAAAGCTCCGGTGAAGAGTCCGCAAGAGGCGAGACTCTCTCTCG 2693
Db 2614 ACAGGCTCCGGGAAAAAGAGGCGGGTGAAGCACTCTCTGGAGGCGAGACTCTCTCTCG 2673
QY 2694 GGCATCGGCAAGACAGGCGCAGAGCCCGCTAAAGAGACTCAATTTGGTCAAGTGGC 2753
Db 2674 GGAACCGGAAAGCGGGCCAGCAGCTGCAAGAAAAAGATTGATTTGGTCAAGTGA 2733
QY 2754 GACTCAGAGTCAAGTCCCGGATTCACAACTCTGGAGAAAGCTCAGACACCCCGGCTGCT 2813
Db 2734 GAGCGAGACTCAAGTACCTGACCCCGAGCTCTGAGACAGCACACAGCCCTCTGGT 2793
QY 2814 GTGGGACCTACTCAATAGGCTTCAGGCGGTGGCGCACAATGGCAGACATAACGAAGC 2873
Db 2794 CTGGGAACCTAATCAATGGCTTACAGGCAAGTGGCGCACATGGCAGACATAACGAAGC 2853
QY 2874 GCGGACGAGTGGGTAAATGCTCAGAAATTTGCAATTTGCGATTCCATGCTGGGCGAC 2933
Db 2854 GCGGACGAGTGGGTAAATTTCTCGGAAATTTGGCATTTGCGATTCCATGATGGGCGAC 2913
QY 2934 AAGATATCAACCAACGACACCCGCACTTGGGCTTGGCCCACTTAATAACCACTCTAC 2993
Db 2914 AAGATATCAACCAACGACACCCGCACTTGGGCTTGGCCCACTTAATAACCACTCTAC 2973
QY 2994 AAGCAATCTCCAGTCTTCAACGAGGGGGCAGAGCAAGCAACCACTACTTGGCTACAG 3053
Db 2974 AAGCAATCTCCAGTCTTCAACGAGGGGGCAGAGCAAGCAACCACTACTTGGCTACAG 3030
QY 3054 ACCCGCTGGGGATTTTGAATTTCAACAGATTCACCTGCCATTTTCAACGCTGAG 3113
Db 3034 ACCCGCTGGGGATTTTGAATTTCAACAGATTCACCTGCCATTTTCAACGCTGAG 3090
QY 3114 CAGGACCTATCAACCAATTTGGGATTTCCGGGCCAAGAGACTCAACTTCAACTCTTC 3173
Db 3094 CAGGACCTATCAACCAATTTGGGATTTCCGGGCCAAGAGACTCAACTTCAACTCTTC 3150
QY 3174 AACATCAAGTCAAGAGGCTCAGACGAATGATGGCTCAACCACTGCTAATTAACCTT 3233
Db 3154 AACATCAAGTCAAGAGGCTCAGACGAATGATGGCTCAACCACTGCTAATTAACCTT 3210
QY 3234 ACCAGACGAGTTCAGTCTTCTCGAATCTGGAGTACCAAGCTTCCGATCGTCTCGCTCT 3293
Db 3214 ACCAGACGAGTTCAGTCTTCTCGAATCTGGAGTACCAAGCTTCCGATCGTCTCGCTCT 3270
QY 3294 GCGCACAGAGGCTGCTCTCTCGGTTCCGGCGGAGGTGTTCAATTCGCGCAATACGCG 3353
Db 3274 GCGCACAGAGGCTGCTCTCTCGGTTCCGGCGGAGGTGTTCAATTCGCGCAATACGCG 3330
QY 3354 TACCTGACGCTCAACATGGCAGCAGAGCGTGGGAGCTTCACTCTTTTACGCTGGA 3413
Db 3334 TACCTGACGCTCAACATGGCAGCAGAGCGTGGGAGCTTCACTCTTTTACGCTGGA 3390
QY 3414 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACACTTTTACCTTCACTCACTTGA 3473
Db 3394 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACACTTTTACCTTCACTCACTTGA 3450
QY 3474 GAAATGCTTTTCAACAGAGCTACGGGCAACAGGCAAGGCTGGAACGGGCTGATGATCT 3533
Db 3454 GAAATGCTTTTCAACAGAGCTACGGGCAACAGGCAAGGCTGGAACGGGCTGATGATCT 3510
QY 3534 CTGATGACCAATATCTGATTAACCTGAAACAGAACTCAATATGATGCTGGGCAAC 3593
Db 3514 CTGATGACCAATATCTGATTAACCTGAAACAGAACTCAATATGATGCTGGGCAAC 3570
QY 3594 AACAGAGCTTGTGTTTACGCTGGGTCTCAGCTGGCATGCTGTTTCAAGCCCAAAAC 3653
Db 3574 CAGTCAAGGCTTCAAGTCTTCTCAGGCGGAGCGAGTACATTCGGGACCAAGTCTAAGAAC 3630

QY 3654 TGGCTACCTGAGCCCTGTTATCGGAGCAGCGGTTTCTTAAACAAAAACAGACCAAC 3713
Db 3634 TGGCTTCTCGAGCCCTGTTATCGGAGCAGCGGTTTCTTAAACAAAAACAGACCAAC 3690
QY 3714 AACAGCAATTTTACCTGAGCTGGTGTCTTCAAAATATTAACCTCAATGGGCGTGAATCATC 3773
Db 3694 AACAGCAATTTTACCTGAGCTGGTGTCTTCAAAATATTAACCTCAATGGGCGTGAATCATC 3750
QY 3774 ATCAACCTTGGCAGCTGATATGGCTTCAACAAAGAGAGAGAGCAAGTCTTTCCCATG 3833
Db 3754 ATCAACCTTGGCAGCTGATATGGCTTCAACAAAGAGAGAGAGAGCAAGTCTTTCCCATG 3810
QY 3834 AGCGGTGATGATTTTGAAGAAAGAGCGCGGAGCTTCAACACTGATTTGACCAAT 3893
Db 3814 AGCGGTGATGATTTTGAAGAAAGAGCGCGGAGCTTCAACACTGATTTGACCAAT 3870
QY 3894 GTCATGATTTACAGACGAGAGGAAATTAAGCACTTAACCTGTGGCCACCGAAGATTT 3953
Db 3874 GTCATGATTTACAGACGAGAGGAAATTAAGCACTTAACCTGTGGCCACCGAAGATTT 3930
QY 3954 GGGACCGTGGCAGTCAATTTTCCAGAGCAGACAGACCACTTGGCCAGCGAGATGCTAT 4013
Db 3934 GGGACCGTGGCAGTCAATTTTCCAGAGCAGACAGACCACTTGGCCAGCGAGATGCTAT 3990
QY 4014 GCTATGGGAGCAATTAACCTGGCAATGGTGTGGCAAGATAGAGAGGTGATCTGACGGTCCC 4073
Db 3994 GCTATGGGAGCAATTAACCTGGCAATGGTGTGGCAAGATAGAGAGGTGATCTGACGGTCCC 4050
QY 4074 ATTTGGGCGCAAAATTTCTTCAACAGATGAGCACTTTCACCCGCTCTCTTATAGGCGGCG 4133
Db 4054 ATTTGGGCGCAAAATTTCTTCAACAGATGAGCACTTTCACCCGCTCTCTTATAGGCGGCG 4110
QY 4134 TTTGACCTCAAGAACCGGCTCTCTGATCTGATCAATAAAACAGCTGTTCTCTGGAAT 4193
Db 4114 TTTGACCTCAAGAACCGGCTCTCTGATCTGATCAATAAAACAGCTGTTCTCTGGAAT 4170
QY 4194 CCGCGGCGGAGTTTCACTCAAAAGTTGGCTTCAATGATCAACCCATGATCTCAACAGA 4253
Db 4174 CCGCGGCGGAGTTTCACTCAAAAGTTGGCTTCAATGATCAACCCATGATCTCAACAGA 4230
QY 4254 CAAATGATGTGGAAATTTGAATGGAGCTGCGAGAAAGAAACAGCAAGGCTGGAATCCC 4313
Db 4234 CAAATGATGTGGAAATTTGAATGGAGCTGCGAGAAAGAAACAGCAAGGCTGGAATCCC 4290
QY 4314 GAAATGCAATCAATCAATTAATGCAAAATCTCCCAAGCTTATTTACTGTGCAAC 4373
Db 4294 GAAATGCAATCAATCAATTAATGCAAAATCTCCCAAGCTTATTTACTGTGCAAC 4350
QY 4374 AATGACCTTATACGAGGCTGCGCCCATTTGGCACCCGTTACCTTACCCGCTCCCTGTA 4433
Db 4354 AATGACCTTATACGAGGCTGCGCCCATTTGGCACCCGTTACCTTACCCGCTCCCTGTA 4410
QY 4434 TTAAGTGAATTAATCAATTAACCGGTTGATGTTTCAAGTGAACCTTGTCTCTGCTT 4493
Db 4414 TTAAGTGAATTAATCAATTAACCGGTTGATGTTTCAAGTGAACCTTGTCTCTGCTT 4470
QY 4494 CTTATCTTATCGGTTACATGTTATAGCTTACATTAACCTGTTGCGCTTCGCG 4553
Db 4474 CTTATCTTATCGGTTATCAATGTTTCAATGTTTACATTAACCTGTTTATCAATTAAC 4530
QY 4554 ATTAAGACTTACGATGAGGTTTACCCCTATGATGAGGTTGCGCACCTGCTGCG 4613
Db 4534 ATTAAGACTTACGATGAGGTTTACCCCTATGATGAGGTTGCGCACCTGCTGCG 4574
QY 4614 CGCTCGCTGCTCGGTGGGCGCTGCGGCAAAAGTCCGAGACGCGCAGAGGCTGCTCT 4673
Db 4574 CGCTCGCTGCTCGGTGGGCGCTGCGGCAAAAGTCCGAGACGCGGCGGCTGCTCT 4634
QY 4674 GCGGCGCCACCGAGCGAGCGCGCAGAGAGGAGTGGGCA 4718
Db 4634 GCGGCGCTCAGTGAAGGAGCGAGCGCGCAGAGAGGAGTGGGCA 4679

RESULT 2
US-08-254-358-1
Sequence 1, Application US/08254358
Patent No. 5658785
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,358
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5658785and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-254-358-1
Query Match 65.3%; Score 3082.4; DB 1; Length 4680;
Best Local Similarity 80.0%; Pred. No. 0;
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

343 TGATTAAGTCCCGAGGACCTTGACGGGACCTGCGGACCTTCTGACGCTTGTGA 402
417 GCTGGTGGCCGAGAGGAATGGAGGTGCCCCGGAATTCGACATGATCTGAATCTGA 476
403 ACTGGGTGGCCGAGAGGAATGGAGGTGCCCCGGAATTCGACATGATCTGAATCTGA 462
477 TTGAGCAGGACCCCTGACCGTGGCCGAGAGGTGACCGGACCTTCTGATCCAAATGCG 536
463 TTGAGCAGGACCCCTGACCGTGGCCGAGAGGTGACCGGACCTTCTGATCCAAATGCG 522
537 GCGCGGTGAGTAAGCCCGGAGGCGCTCTTTTGTTCAGTTCGAGAAAGGGAGTCTT 596
523 GCGGTGAGTAAGGCGCCGAGGCGCTCTTTTGTTCAGTTCGAGAAAGGGAGTCTT 582
597 ACTTCCACCTCCATTTCTGTTGAGAGCAAGGGGGGTCAATCATGATGCTGGGCGCT 656
583 ACTTCCACATGACGTGCTGTGGAACCAAGGGGGGTCAATCATGATGCTGGGCGCT 642
657 TCCGAGTCAGATTTAGGAGCAAGCTGATGAGCAATCTACCGCGGATCGAGCCGACCC 716
643 TCCGAGTCAGATTTAGGAGCAAGCTGATGAGCAATCTACCGCGGATCGAGCCGACCT 702
717 TGCCCAACTGCTTTCGGGTGACCAAGACGCTATGAGCCCGGAGGGGGAACAAGTGG 776
703 TGCCCAACTGCTTTCGGGTGACCAAGACGCTATGAGCCCGGAGGGGGAACAAGTGG 762
777 TGAGCAGGTCTCATCTCCCACTACCTCTGCGCAAGATCTGACCCGAGCTCAATGAGG 836
763 TGAGTGAATCTCATCTCCCACTACCTCTGCGCAAGATCTGACCCGAGCTCAATGAGG 822
837 CGTGAATCATGAGAGATATATAGCGCTGTTTGAACCTGCGCGAGCGCAACCGC 896
823 CGTGAATCATGAGAGATATATAGCGCTGTTTGAACCTGCGCGAGCGCAACCGC 882
897 TCGTGGCGGACGACCTGACCCAGTCAAGCCAGAGCGAGAACAGAGAAATCTGA 956
883 TGTGTGGCGGACGACCTGACCCAGTCAAGCCAGAGCGAGAACAGAGAAATCTGA 942
957 ACCCAATTTGACGCGCGCTGATCCGATCAAAACCTCGCGCGCTACATGAGAGCTGG 1016
943 ATCCCAATTTGATGCGCGCGCTGATCCGATCAAAACCTCGCGCGCTACATGAGAGCTGG 1002
1017 TCGGAGTGGCTGAGACCGGGGAGCATCACTCCGAGAAAGATGATCCAGAGAGACCAAG 1076
1003 TCGGAGTGGCTGAGACCGGGGAGCATCACTCCGAGAAAGATGATCCAGAGAGACCAAG 1062
1077 CTTGATCATCTCTTCAAGCGCGCTTCCAACTCGGGTCCGAGATCAAGCCGCTCTGG 1136
1063 CTTGATCATCTCTTCAAGCGCGCTTCCAACTCGGGTCCGAGATCAAGCCGCTCTGG 1122
1137 ACAATGCGGCAAGATCATGAGCGCTGACCAATCCGCGCGCGCATCTGATGAGCGCCG 1196
1123 ACAATGCGGCAAGATCATGAGCGCTGACCAATCCGCGCGCGCATCTGATGAGCGCCG 1182
1197 CTCGCGCGCGGACATTTAAACCAACCGCATCTACCGCATCTGAGAGCTGAACGCTACG 1256
1183 AGCCCGTGAAGACATTTCAAGCAATCGATTTATTAATTTTGAACCTAAACGGGTACG 1242
1257 AACCTGCTACGCGCGCTCTTCTTCTGAGCTGGGCGCCGAGAAAGGTTGGGAAGCGCA 1316
1243 ATCCCAATTTGAGGCTCTTCTTCTTCTGAGAGGCGCACGAAAGGTTGGGAAGCGCA 1302
1317 ACAACATCTGAGCTGTTTGGGCGCGGACCAACCGGAGCAACAACATCTGCGAGAGCATCG 1376
1303 ACAACATCTGAGCTGTTTGGGCGCGGACCAACCGGAGCAACAACATCTGCGAGAGCATCG 1362
1377 CCGACGCGGTGCCCTTCTACGCGTGCCTCAATGAGCAATGAGAACTTTCCTTCAAG 1436
1363 CCGACGCGGTGCCCTTCTACGCGTGCCTCAATGAGCAATGAGAACTTTCCTTCAAG 1422
1437 ATTGCGTCAACAAGATGATCTGTTGGAGAGAGGCAAGATGACGCGCAAGGTCTGG 1496
1423 ACTGTGCGACAAGATGATGATCTGTTGGAGAGAGGCAAGATGACGCGCAAGGTCTGG 1482

1497 AGTCCGCAAGGCGCATTTCTGCGGCGAGCAAGGTCGCGTGA CCAAAGTGCAGTGGT 1556
1483 AGTCCGCAAGGCGCATTTCTGCGGCGAGCAAGGTCGCGTGA CCAAAGTGCAGTGGT 1542
1557 CCGCCAGATGACCCCAACCCCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1616
1543 CCGCCAGATGACCCCAACCCCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1602
1617 TTGACGGGAACAGACCACTTTCAGACCAAGACCGGTCGAGACCGGATGTTCAAT 1676
1603 TTGACGGGAACAGACCACTTTCAGACCAAGACCGGTCGAGACCGGATGTTCAAT 1662
1677 TTGACCTACCCGCGCGTTCGAGACATGACTTTGGCAAGGTCAGCAAGGAGGTAAG 1736
1663 TTGACCTACCCGCGCGTTCGAGACATGACTTTGGCAAGGTCAGCAAGGAGGTAAG 1722
1737 AGTTCTTCGCGTGGCGCAGATCAAGTACCGAGGTCGCGATGATGTTCAAGTGA 1796
1723 ACTTTTTCGCGTGGCGCAGATCAAGTACCGAGGTCGCGATGATGTTCAAGTGA 1782
1797 AGGTTGAGGCAACAAAGACCCCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1856
1783 AGGTTGAGGCAACAAAGACCCCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1842
1857 CCGCCGCGTTCGCGTGGCGCAGATCAAGTACCGAGGTCGCGATGATGTTCAAGTGA 1916
1843 TGGCGAGTCAAGTGGCGCAGATCAAGTACCGAGGTCGCGATGATGTTCAAGTGA 1899
1917 CCGACAGATCAACAAAGACCCCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1976
1900 CAGACAGATCAACAAAGACCCCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 1959
1977 GCAAGACATGCGGAGAAATGAAATTCAGATTTCAACATTTGCTTCAAGCAAGGAGGAG 2036
1960 GCAAGACATGCGGAGAAATGAAATTCAGATTTCAACATTTGCTTCAAGCAAGGAGGAG 2019
2037 ACTGTCAGAGTGGTCCCGCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 2093
2020 ACTGTCAGAGTGGTCCCGCGCGGTGATGCTCACTTCAACCAACATGTCGCGTGA 2076
2094 CCGTACGAAATCTGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCGTC 2153
2077 CCGTACGAAATCTGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCGTC 2133
2154 CCGCTCGCATCTGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCGTC 2213
2134 CCGCTCGCATCTGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCGTC 2193
2214 AAACGAGTATGCTGCGATGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCG 2273
2194 AAACGAGTATGCTGCGATGTCGATTCATCATCTGTCGAGGCGGTCGCGATGTCGTCG 2253
2274 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2333
2254 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2313
2334 AAGCAGAGAGACCGCGCGGTGTCGTCGTCCTTCTGCTCAAGTACTCGGACCTTCAAC 2393
2314 CATTAAGAGAGACCGCGGTGTCGTCGTCCTTCTGCTCAAGTACTCGGACCTTCAAC 2373
2394 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2453
2374 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2433
2454 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2513
2434 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2493
2514 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2573
2494 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2553

2574 GTCTTCAGGCGCAAGAGCGGTTCTGAACTCTGCTGCTGAGGAGGAGCGCTAAG 2633
2554 GTCTTCAGGCGCAAGAGCGGTTCTGAACTCTGCTGCTGAGGAGGAGCGCTAAG 2613
2634 ACCGCTCTGGAAGAAAGTCCGCTGAGAGAGTGCACAAAGGCAAGCTTCTCTCG 2693
2614 ACCGCTCTGGAAGAAAGTCCGCTGAGAGAGTGCACAAAGGCAAGCTTCTCTCG 2673
2694 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2753
2674 GGCATTCGCGAGTGTGAGCTTGAACCTGAGACCCCGGAGCCCAAGCCCAACAGCA 2733
2754 GACTCAGATGAGTCCCGATCCCAACACTCTGAGAGAACTTCAGCAACCCCGCTGCT 2813
2734 GAGCGAGTCTGAGTACTGACCCCGAGCTTCTGAGAGAGCAACAGGAGGAGGAG 2793
2814 GTGGAGCTTCAATCAATGAGCTTCAAGCGGTGCGACCAATGCGAGCAATTAAGAGG 2873
2794 GTGGAGCTTCAATCAATGAGCTTCAAGCGGTGCGACCAATGCGAGCAATTAAGAGG 2853
2874 GCGAGGAGTGGGTAAAGCTTCAGAAATTTGGCATTTGGATTCGATGCTGCGGAG 2933
2854 GCGAGGAGTGGGTAAAGCTTCAGAAATTTGGCATTTGGATTCGATGCTGCGGAG 2913
2934 AGAGTATCAACCAAGAGCCGACCTGCGGCTTCCAGCTTCAATTAACCACTCTAC 2993
2914 AGAGTATCAACCAAGAGCCGACCTGCGGCTTCCAGCTTCAATTAACCACTCTAC 2973
2994 AAGCAATCTCAAGTCTTCAAGCGGAGCGAGCAAGCAACCACTTCTGCGTAC 3053
2974 AAGCAATCTCAAGTCTTCAAGCGGAGCGAGCAAGCAACCACTTCTGCGTAC 3030
3054 ACCGCTCTGGAAGAAAGTCCGCTGAGAGAGTGCACAAAGGCAAGCTTCTCTCG 3113
3031 ACCGCTCTGGAAGAAAGTCCGCTGAGAGAGTGCACAAAGGCAAGCTTCTCTCG 3090
3114 CAGCGATCATCAACCAATGAGGAGTTCGCGGCGCAAGAGACTCAACTTCAACTCTTC 3173
3091 CAGCGATCATCAACCAATGAGGAGTTCGCGGCGCAAGAGACTCAACTTCAACTCTTC 3150
3174 AACATCAAGTCAAGAGGTCAGAGCAAGTATGAGGTCAGCAACCACTGCTAATACCTT 3233
3151 AACATCAAGTCAAGAGGTCAGAGCAAGTATGAGGTCAGCAACCACTGCTAATACCTT 3210
3234 ACCAGAGGTCAGAGTCTTCTGAGCTGAGAGTACCACTTCCGATGCTGCTGCTGCT 3293
3211 ACCAGAGGTCAGAGTCTTCTGAGCTGAGAGTACCACTTCCGATGCTGCTGCTGCT 3270
3294 GCGACCAAGGAGTCTTCTGAGCTGAGAGTACCACTTCCGATGCTGCTGCTGCTGCT 3353
3271 GCGACCAAGGAGTCTTCTGAGCTGAGAGTACCACTTCCGATGCTGCTGCTGCTGCT 3330
3354 TACCTGACGCTCAACCAATGAGGTCAGAGCAAGGTCGAGGAGTTCATCTTACTGCTGAA 3413
3331 TACCTGACGCTCAACCAATGAGGTCAGAGCAAGGTCGAGGAGTTCATCTTACTGCTGAA 3390
3414 TATTTCTCTTCTGAGTGTGAGAGAGGCAACATTTACTTCTGAGTCAACCTTTGAG 3473
3391 TATTTCTCTTCTGAGTGTGAGAGAGGCAACATTTACTTCTGAGTCAACCTTTGAG 3450
3474 GAGGTCCTTCTCAAGAGTGTGAGAGGCAACATTTACTTCTGAGTCAACCTTTGAG 3533
3451 GAGGTCCTTCTCAAGAGTGTGAGAGGCAACATTTACTTCTGAGTCAACCTTTGAG 3510
3534 CTCATGACCAATACCTGATTAATCTGAAACAGAACTCAAAATCAAGTCCGAAAGTCCAA 3593
3511 CTCATGACCAATACCTGATTAATCTGAAACAGAACTCAAAATCAAGTCCGAAAGTCCAA 3570
3594 AACAGAGTCTGCTGTTGAGCGGTGCTCAAGCTGAGTGTCTTCAAGCCCAAAAC 3653
3571 CAGTCAAGGCTTCAAGTGTCTCAAGGCGGAGGAGTGAATTCGGAACAGTCTTAAGAAC 3630
3654 TGGCTACCTGAGCCCTGTTATGCGAGAGCGGTTCTTAACCAAAACAGCAACAAAC 3713

Db	3631	TGGCTCTCTGSAACCTGTATCCGCGCAGAGGAGTATCAAGAGCATCTGGCGATATACAC	3690
Qy	3714	AACAGCAATTTTACTGTGACCTGTGCTTCAAAATTAATCTCATATGGCGTGAATCATC	3773
Db	3691	AACAGTAAATCTCGTGGACCTGGAGCTACAAAGTACCACTTCATATGGCAGAGACTCTCTG	3750
Qy	3774	ATCAACCTGTGCACTGCTATGGCCCTCACAAAAGCAGCAAGACAAATTCTTTTCCCATTG	3833
Db	3751	GTGAATCCGGGGCCCGCCATGGCAAGCACCAAGAGCATGAAGAAAGTTTTCCTCAG	3810
Qy	3834	AGCGGTGTCATGATTTTGGAAAAAGAGCGCCGAGCTTCAACCTGCATTGGACAT	3893
Db	3811	AGCGGGGTCTTCATCTTTGGGAAGCAAGGCTCAGAAAAAATGTGACATTGAAAG	3870
Qy	3894	GTCAATGATTAACAAGAAAGAAATTAAAGCACTAACCTGTGGCCACGAAAGATT	3953
Db	3871	GTCAATGATTAACAAGAAAGAAATCGAAACACCAATCCCTGTGGCTTACGAGACATAT	3930
Qy	3954	GGGACCGTGGAGTCAATTTTCCAGAGCAGACAGACCCCTGGCAGCCGAGATGTGCAT	4013
Db	3931	GGTTCGTATCTACCAACCTCCAGAGAGCAACAGACAGACTTACCGCAGATGTCAAC	3990
Qy	4014	GCTATGGAGCATTAACCTGGCATGCTGTGGCAAGATAGAGCTGTATCTGCAAGGTGCC	4073
Db	3991	ACACAGGCGTTCCTTCAGGCAAGTCTGGCAGCAAGATGTATCTTCAGGGGCCC	4050
Qy	4074	ATTTGGGCAAAATTCCTCACAAGATGAGACATTTCAACCCGCTCTCTTATGGCGGC	4133
Db	4051	ATCTGGGCAAAAGATTCCACACACGAGCGGACATTTTCAACCCCTCTCATGGGTGA	4110
Qy	4134	TTTGGACTCAGAAACCGGCTCTCTCAATCTCTATCAAAAACAGGCTGTCTCTGCCAAT	4193
Db	4111	TTTGGACTTAAACACCTCTCTCCACAGATTTCTATCAAGAAACACCCGCTGACTGGCAAT	4170
Qy	4194	CCTCGGCGGAGTTTTCAGTACAAAGTTTCTTCATTCATCAACCACTATCTCACAGGA	4253
Db	4171	CCTTGACACACTTCAGTGGGGCAAAAGTTTGCTTCTTCATCAACAGTACTCCACGGGA	4230
Qy	4254	CA-AGTAGTGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGATCC	4312
Db	4231	CACGGTCAGCGTGGAGATCGATGGGAGCTCAGAAAGAAACAGCAACCTGGAAATCC	4290
Qy	4313	CGAAGTCAGTACATCAATTAATGCAAAATCTGCCAAGTGAATTTTACTGTGACAA	4372
Db	4291	CGAATTCAGTACACTTCCAACTACAAAGTCTTTAATGTGGACTTACCGTGGATAC	4350
Qy	4373	CAATGACTTATACTGAGCTCGCCCATTTGGCAACCCGTTACTTACCGTCCCTGTA	4432
Db	4351	TAAATGGCGTATTCAGAGCCTCGCCCATTTGGCACAGATACCTGAATCTGTGA	4410
Qy	4433	ATTAGGTATATCAATTAACCGGTTGATTTGTTCACTTGAACCTTGGTCTCGTCT	4492
Db	4411	ATTGCTGTATTAATCAATTAACCGTTTATTTGTTGCAATTAATCTTGTGCTGAT	4470
Qy	4493	TCTTATCTTATCGGTTACATGGTATTAAGCTTACATTAATCTGTTGGTTCGCTTCGC	4552
Db	4471	TTCTTTCTTATCTAGTTTCCATGGCTAGTGAATTAATAGATGGCGGTTAATCATTAA	4530
Qy	4553	GATAAAGACTTACGTCAATCGGGTTAACCCCTAGTGAATGGAAATTCCTCCCTCTGCG	4612
Db	4531	CTACAGGA-----ACCCCTAGTGAATGGAAATTCCTCCCTCTGCG	4574
Qy	4613	GCCTCGCTCGCTCGTGGGGCTTGGCAACCAAGTTCGCAACGCGAGACTCTGCTC	4672
Db	4575	GCCTCGCTCGCTCACTGAGGCGGGGCGACCAAAAGTGGCCCAAGCCCGGGCTTTGCC	4634
Qy	4673	TGCGGCGCCCAACGAGCGAGCGAGCGCCCAAGAGGAGTGGCCAA 4718	
Db	4635	GGGGCGGCTCTAGTAGCGAGGAGCGCCCAAGAGGAGTGGCCAA 4680	

RESULT 3

US-08	175-391-1	Application US/08475391
Sequence 1	1	
Patent No.	5786211	
GENERAL INFORMATION:		
APPLICANT:	Johnson, Philip R.	
TITLE OF INVENTION:	Adeno-Associated Virus Materials and	
TITLE OF INVENTION:	Methods	
NUMBER OF SEQUENCES:	3	
CORRESPONDENCE ADDRESS:		
ADDRESS:	Marshall, O'Toole, Gerstein, Murray & Borun	
STREET:	6500 Sears Tower, 233 S. Wacker Drive	
CITY:	Chicago	
STATE:	Illinois	
COUNTRY:	USA	
ZIP:	60606	
COMPUTER READABLE FORM:		
MEDIUM TYPE:	Floppy disk	
COMPUTER:	IBM PC compatible	
OPERATING SYSTEM:	PC-DOS/MS-DOS	
SOFTWARE:	Patentin Release #1.0, Version #1.25	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER:	US/08/475,391	
FILING DATE:	07-JUN-1995	
CLASSIFICATION:	435	
PRIOR APPLICATION DATA:		
APPLICATION NUMBER:	08/254,358	
FILING DATE:		
ATTORNEY/AGENT INFORMATION:		
NAME:	No. 5786211and, Greta E.	
REGISTRATION NUMBER:	35,302	
REFERENCE/DOCKET NUMBER:	31975	
TELECOMMUNICATION INFORMATION:		
TELEPHONE:	(312) 474-6300	
TELEFAX:	(312) 474-0448	
TELEX:	25-3856	
INFORMATION FOR SEQ ID NO: 1:		
SEQUENCE CHARACTERISTICS:		
LENGTH:	4680 base pairs	
TYPE:	nucleic acid	
STRANDEDNESS:	single	
TOPOLOGY:	linear	
MOLECULE TYPE:	DNA (genomic)	
US-08	475-391-1	
Query Match	65.3%	Score 3082.4; DB 1; Length 4680;
Best Local Similarity	80.0%	Pred. No. 0;
Matches 3781;	Conservative 0;	Mismatches 891; Indels 54; Gaps 11.
QY	1	TTGGCCCACTCCCTCTGTCGCGCTCGCTCGGTGGGGGCTCGCGACCAAGCTCCGC 60
DB	1	TTGGCCCACTCCCTCTGTCGCGCTCGCTCGGTGAGGCGCGGCGACCAAGGTCCGC 60
QY	61	AGACGGCAGAGCTCTGCTCTGCGCGGCCCACTCGACGCGAGCGAGCGCGAGAGGAGTG 120
DB	61	CGAGCGCCCGGGCTTTGGCCCGGGCGGCTCACTGAGCGAGCGCGCGAGAGGAGTG 120
QY	121	GGCAACTCCATCATTAAGGGGTAACTCGGAGAGCGGCTCCACAGCTGCGCGGATCGCGTGA 180
DB	121	GCCAACTCCATCATTAAGGGGT-----TCTTGAGAGGGGTGAGATCGTGA 163
QY	181	CGTAATTAAGCTCAATAGG--GAGTGTCTGTATTAATTAAGCTGTCACTGAGTCTTTTGC 237
DB	164	CGTAATTAAGCTCAATAGGTTAAGGAGGTCTGTATTAAGAGTCACTGAGTG--TTTTC 222
QY	238	GACATTTTGGCAGCAGCGTGGCCATTTAAGGTATATATGCGCGAGTAGAGGAGGAT 297
DB	223	GACATTTTGGCAGCAGCGTGGTCACTGAGGTATTTAAGCCGAGTAGAGCAGAGGT 282
QY	298	CTCATTTTGG--ACGCGCAAAATTTGAAGAGCAGAGCGAGCGATCCGCGGCTTTCAAGATCG 356
DB	283	CTCATTTTGAAGGGGAGGTTTGAACGCGAGCGCGCATCCGCGGTTTACAGATTTG 342
QY	357	TGATCAAGTGCAGAGCGAAGCTGAGCAGACCTTCCGCGGCAATTTTCACTTCGTTTGA 416

Db 343 TGAATAGATCCCGAGGACCTTGACCGGGATCTGCCCGCATTTTGACAGCGTTTGA 402
Qy 417 GCTGGGTGGCCGAGAAAGAAATGGAGCTGCCCCGGGATTTGACATGATCTGAATCTGA 476
Db 403 ACTGGGTGGCCGAGAAAGAAATGGAGCTGCCCCGGGATTTGACATGATCTGAATCTGA 462
Qy 477 TTGAGCAGGCAACCCCTGACCGGTGGCCGAGAAATGGAGCTGCCCCGGGATTTGACATGATCTGAATCTGA 536
Db 463 TTGAGCAGGCAACCCCTGACCGGTGGCCGAGAAATGGAGCTGCCCCGGGATTTGACATGATCTGAATCTGA 522
Qy 537 GCCCGGTGAGTAAAGCCCGGAGGCCCTCTTTTGTTCAGTTTGAAGAGGCGAGTCT 596
Db 523 GCCGTGAGTAAAGCCCGGAGGCCCTCTTTTGTTCAGTTTGAAGAGGCGAGTCT 582
Qy 597 ACTTCCACCTTCATATTTGTGTGAGAACACGCGGGGTCAATTCATGTGTGTGGCCGT 656
Db 583 ACTTCCACATGACAGTCTGTGTGAAACACCGGGGTGAAATTCATGTGTGTGGGACGTT 642
Qy 657 TCCGATGATGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGGATGAGCCGACCC 716
Db 643 TCCGATGATGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGGATGAGCCGACCTT 702
Qy 717 TGCCCAACTGGTTCGCGGTGACCAAGACGCGTAAATGGCGCGAGGGGAAACAAGTGG 776
Db 703 TGCCCAACTGGTTCGCGGTGACCAAGACGCGTAAATGGCGCGAGGGGAAACAAGTGG 762
Qy 777 TGAAGAGTGTATATCCCACTACTCTCTGACCAAGTCAAGCCGAGCTGAGTGG 836
Db 763 TGAATGATGTATATCCCACTACTCTCTGACCAAGTCAAGCCGAGCTGAGTGG 822
Qy 837 CGTGAATATATGAGAGATATTAAGCGCTGTTTGAACCTGGCGAGCGCAAGCGC 896
Db 823 CGTGAATATATGAGAGATATTAAGCGCTGTTTGAACCTGGCGAGCGCAAGCGT 882
Qy 897 TCGTGGCAGCACTGACCCGCTGATCCGGGTCAAAACCTCGCGCGCTACATGAGTGG 956
Db 883 TCGTGGCAGCACTGACCCGCTGATCCGGGTCAAAACCTCGCGCGCTACATGAGTGG 942
Qy 957 ACCCGAATTTGAGCGCGCTGATCCGGGTCAAAACCTCGCGCGCTACATGAGTGG 1016
Db 943 ATCCCAATTTGAGCGCGCTGATCCGGGTCAAAACCTCGCGCGCTACATGAGTGG 1002
Qy 1017 TCGGGTGGCTGAGGACCGGGGATCACTCCCGAGAGAGTGGATCCAGAGAGACAGG 1076
Db 1003 TCGGGTGGCTGAGGACCGGGGATCACTCCCGAGAGAGTGGATCCAGAGAGACAGG 1062
Qy 1077 CCTGTATCATCTCTTCAACGCGCTTCAACTGCGCGTCCAGATCAAGCGCTGG 1136
Db 1063 CCTGTATCATCTCTTCAACGCGCTTCAACTGCGCGTCCAGATCAAGCGCTGG 1122
Qy 1137 ACAATGCGGCAAGATCATGCGCTGACCAATTCGCGCGCTCACTGTTGAGCCCG 1196
Db 1123 ACAATGCGGCAAGATCATGCGCTGACCAATTCGCGCGCTCACTGTTGAGCCCG 1182
Qy 1197 CTCCGCGCGGACATTTAAACCAACCGCATCTACCGCATCTGAGAGTGAAGCGCTAG 1256
Db 1183 AGCCCGTGAAGACATTTCAAGCAATGAGATTTAAATTTTGAACCTAAACCGGCTAG 1242
Qy 1257 AACCTGCTAGCGCGCTCGCTCTTTCTCGGCTGGGCCAGAAAAAGTTGCGGAAGCGCA 1316
Db 1243 ATCCCAATTTAGCGGCTTCCGTCTTTCTGAGATGAGGCAAGAAAAAGTTGCGGAGAGGA 1302
Qy 1317 ACAACCATCTGCTGTTTGGCGCGGCAACAAGGCAAGAACCAATGCGGAGAGCATCG 1376
Db 1303 ACAACCATCTGCTGTTTGGCGCGGCAACAAGGCAAGAACCAATGCGGAGAGCATCG 1362
Qy 1377 CCGACGCGGTGCGCTTACGCGCTGAGCACTGAGCAATGAGAACTTTCCCTTCAATG 1436
Db 1363 CCGACGCGGTGCGCTTACGCGCGTGAAGCACTGAGCAATGAGAACTTTCCCTTCAATG 1422
Qy 1437 ATTGCGTGAACAAGATGATCTGTGTGAGAGGAGCAAGATGACGCGCAAGGTCTGG 1496

Db 1423 ACTGTGCAACAAGATGATCTGTGTGAGAGGAGGAGAAATGATCCGCAAGTGTGG 1482
Qy 1497 AGTCCGCAAGGCAATTTCTGGCGGCAAGGTGGCGGTGAGCAAAAGTCAAGTCT 1556
Db 1483 AGTCCGCAAGGCAATTTCTGGCGGCAAGGTGGCGGTGAGCAAAAGTCAAGTCT 1542
Qy 1557 CCGCCAGATGACACCCCGGTGATGCTCACTCAACCAACATGAGCGCGTGA 1616
Db 1543 CCGCCAGATGACACCCCGGTGATGCTCACTCAACCAACATGAGCGCGTGA 1602
Qy 1617 TTGACGGAACAGACCACTTGAAGACCAAGCGGTTGAGAGCCGATGTTCAAT 1676
Db 1603 TTGACGGAACCTCAACCACTTGAAGACCAAGCGGTTGAGAGCCGATGTTCAAT 1662
Qy 1677 TTGACCTACCCCGCTTGAAGACATGCTTTGGCAAGGTGAACAAGACGAAAGTCAAAG 1736
Db 1663 TTGACCTACCCCGCTTGAAGACATGCTTTGGCAAGGTGAACAAGACGAAAGTCAAAG 1722
Qy 1737 AGTTCTTCGCTGGGCGAGATCAAGTGAACGAGGTGGCGATGATGTTCAAGTCAAG 1796
Db 1723 AGTTCTTCGCTGGGCGAGATCAAGTGAACGAGGTGGCGATGATGTTCAAGTCAAG 1782
Qy 1797 AGGATGAGCCAAACAAAGACCCCGCGATGACGCGATTAAGACGAGCCCAAGCGG 1856
Db 1783 AGGATGAGCCAAACAAAGACCCCGCGATGACGCGATTAAGAGCCCAAGCGG 1842
Qy 1857 CTTGCCCTGATGCGGATCATGACGTGACGCGGAGAGAGTCCGATGACTTG 1916
Db 1843 TGCAGAGTCAAGTGGCGAGCCATGACGTCAGACCGGGA--AGCTTCATGATCACTAG 1899
Qy 1917 CCGAGAGTACCAAAACAAATGTTCTGTCAGCGCGGCAAGTTCAGATGCTGTTCCCT 1976
Db 1900 CAGACAGTACCAAAACAAATGTTCTGTCAGCGGCAAGTTCAGATGCTGTTCCCT 1959
Qy 1977 GCAAGACATGCGAGAGATGATCAGATTTCAATTTGCTTCAAGCAAGGAGAGAG 2036
Db 1960 GCAAGACATGCGAGAGATGATCAGATTTCAATTTGCTTCAAGCAAGGAGAGAG 2019
Qy 2037 ACTGTTCAAGTCTTCCCGCGGTGACAGATTTCAACCGGT--GTCAGAAAGAGGA 2093
Db 2020 ACTGTTCAAGTCTTCCCGCGGTGACAGATTTCAACCGGT--GTCAGAAAGAGGA 2076
Qy 2094 CGATGCGAAACCTGAGCATTCATCATCTGCTGGGGGCGGCTCCGAGATGCTTGT 2153
Db 2077 CGATGCGAAACCTGAGCATTCATCATCTGCTGGGGGCGGCTCCGAGATGCTTGT 2133
Qy 2154 CGGCTGCGATCTGTCAACGTTGACCTGAGTGTGTTTCTGACATTAATGACTT 2213
Db 2134 CTGCTGCGATCTGTCAACGTTGAGTGTGTTTCTGACATTAATGACTT 2193
Qy 2214 AAACAGATGAGCTGCGATGTTTCAAGATTTGCTGAGAGCAACTCTCTGAG 2273
Db 2194 AAATCAAGTATGCTGCTCCGATGTTTCTTCCAGATTTGCTGAGAGCACTCTCTGAA 2253
Qy 2274 GCGATTCGAGAGGTGGGACCTTGAACCTGAGCCCGGAGGCCAAAGCCAGCGAA 2333
Db 2254 GGAATTAAGACAGTGGTGAAGGCTCAACCTGCGCCACACCAACCAAGCCCGAGAGCGG 2313
Qy 2334 AAGCAGAGCAGCGCGGCTGAGTCTTCCGCTGACCAAGTACCTGAGACCTTCAAC 2393
Db 2314 CATTAAGAGCAGCAGGAGGCTTGTGCTTCTGAGTCAAGTACCTGAGACCTTCAAC 2373
Qy 2394 GAGCTGACAGGAGGAGCCCGTCAACGCGCGAGACGACAGCGGCTTCAAGCAGCAAG 2453
Db 2374 GAGCTGACAGGAGGAGCCCGTCAACGCGCGAGACGACAGCGGCTTCAAGCAGCAAA 2433
Qy 2454 GCTTAAGACAGAGCTCAAGAGCGGTGAACAATCGCTACCTGAGTAAACAGCGCGAG 2513
Db 2434 GCTTAAGACAGAGCTCAAGAGCGGTGAACAATCGCTACCTGAGTAAACAGCGCGAG 2493
Qy 2514 GCCGAGTTTCAGAGCGTGTGCAAGAGATGACTTTTGGGGGCAACTCGGCGAGAGA 2573
Db 2494 GCCGAGTTTCAGAGCGCTTAAAGAGATGACTTTTGGGGGCAACTCGGCGAGAGA 2553

2574 GTCTCCAGGCGCAAGAGCGGTTCTCGAAGCTCTCGGTCTGTTGAGAAAGCGCTAAG 2633
2554 GTCTTCCAGGCGGAAAAAGAGGTTCTTGAAGCTCTGGGCTGTGTGAGAACTGTTAAG 2613
2634 ACCGCTCTGTGAAAAAGAGTCCGGGTAGAGCACTGCGCAAGAGCGCAACTCTCTCG 2693
2614 ACCGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCGCAACTCTCTCG 2673
2694 GGCATCGGCAAGACAGGCTCAGACCGCGCTAAAGAGCACTATTTTGTGACACTGGC 2753
2674 GGAACCGGAAAGGCGGCGAGAGCTGCAAGAAAAAGATTGAAATTTGTGCAACTGGA 2733
2754 GACTCAGAGTACGTCCTCCGATCCCAACCTCTCGGAGAACTCTCAGAAACCCCGCTCT 2813
2734 GACCGAAGCTAGTACCTGAGACCCCGAGCTCTCGGACAGCCACAGAGCGCCCTCTGGT 2793
2814 GTGGAGCTTACTAGCAATGGCTTCAAGCGGTGCGGCACTCAATGCGACAAATACGAAAGC 2873
2794 CTGGGAATTAATAGATGGCTACAGGCACTGGCGCACTCAATGCGACAAATACGAAAGC 2853
2874 GCCGACGAGTGGGTAAATGCTTCAAGAAATTTGGCAATTTGCAATGTGCTGGGCGAC 2933
2854 GCCGACGAGTGGGTAAATCTCCGGAATTTGGCAATTTGCAATGTGCTGGGCGAC 2913
2934 AGAGTCATCAACCAAGAGCGGCACTGGGCTTGGCCCACTCAATTAACCACTCTAC 2993
2914 AGAGTCATCAACCAAGAGCGGCACTGGGCTTGGCCCACTCAATTAACCACTCTAC 2973
2994 AAGCAATCTCAGTGTCTTCAACGGGGCGGCAAGCAACCACTACTTCTGGCTACAGC 3053
2974 AAGCAATCTCAGTGTCTTCAACGGGGCGGCAAGCAACCACTACTTCTGGCTACAGC 3030
3054 ACCCGCTGGGGGTATTTGATTTCAAGATTTCACTGGCACTTTTCAAGAGTACG 3113
3031 ACCCGTGGGGGTATTTGATTTCAAGATTTCACTGGCACTTTTCAAGAGTACG 3090
3114 CAGGACTCATCAACCAATTTGGGATTTCCGCGCCAGAGACTCAATTTCAATCTCTC 3173
3091 CAAAGACTCATCAACCAATTTGGGATTTCCGCGCCAGAGACTCAATTTCAATCTCTC 3150
3174 AAGATCCAGTCAAGAGAGTCAAGCAAGATAGCGGTCAACACATCTGCTAATACCTT 3233
3151 AAGATCCAGTCAAGAGAGTCAAGCAAGATAGCGGTCAACACATCTGCTAATACCTT 3210
3234 ACCGACAGTTCAGATCTTCTCGGACCTCGAGATACAGCTTCCGTAGCTCTCGGCTCT 3293
3211 ACCGACAGTTCAGATCTTCTCGGACCTCGAGATACAGCTTCCGTAGCTCTCGGCTCT 3270
3294 GCGACAGAGGCTGCTCTCCCTCCGTTCCGCGCGAGCTGTTCATGATTTCCGCAATACGCG 3353
3271 GCGACAGAGGATGCTCTCCCGCTTCCGCGAGAGCTTCTCATGTGTCCCAAGATAGGA 3330
3354 TACCTGAGCTCAACCAATAGGAGCGCAAGCCGTGGAGCTTCACTCTTACTGCTGGA 3413
3331 TACCTGAGCTCAACCAATAGGAGCTCAAGCAGTAGGAGCTCTTCACTTACTGCTGGA 3390
3414 TATTTCCCTTTCAGATGCTGAGAAAGCGGCAACCTTACTTACTGAGTACCTTTGAG 3473
3391 TACTTTCTTTCAGATGCTGAGTACCGGAAACCTTACTTACTGAGTACCTTTGAG 3450
3474 GAAAGTCTTTCAGAGAGCTACGCGACAGCCAGAGCTGAGCCGCGTATGATCT 3533
3451 GAGCTTCTTTCAGAGAGCTACGCGACAGCCAGAGCTGAGCCGCGTATGATCT 3510
3534 CTCATCGACCAATACCTGTATTAATCTGAAAGAACTCAAAATCACTCCGGAAGTCCCA 3593
3511 CTCATCGACCAATACCTGTATTAATCTGAGCAAGAACTCAAAAGTCAAGCAACCG 3570
3594 AACAGAGACTGTGTATAGCGGTGCTCTCAGCTGGCAATGTCTGTTCAGCCCAAAAC 3653
3571 CAGTCAAGCTTCAAGTCTTCTCAAGCGGAGGAGTGAATTTGGGACCAAGTCTAGAAC 3630

3654 TGCTACCTTGAGACCTGTATTCGCGAGAGCGGCTTTCTPAAACAAAACAGACAAAC 3713
3631 TGCTTCTTGAGACCTGTATTCGCGAGAGCGAGTATCAAAAGCATCTGCGTAAACAC 3690
3714 AACAGCAATTTTACCTGAGCTGTGTCTCAAAATATTAACCTCAATGGGGTGAATCCATC 3773
3691 AACAGCAATTTTACCTGAGCTGTGTCTCAAAATATTAACCTCAATGGGGTGAATCCATC 3750
3774 ATCAACCTTGAGACCTGTATTCGCGAGAGCGGCTTTCTPAAACAAAACAGACAAAC 3833
3751 GTGAATTCGCGGCGCGCGACATGCAACGCACAGAGAGTGAAGAAAGTTTTTCCGAG 3810
3834 AGCGGTGTATGATTTTGGAAAAAGAGCGCGGAGCTTCAAAACCTGATTTGACAT 3893
3811 AGCGGGTGTCTACTTTTGGGAAAGCAAGGCTCAGAGAAACAAATGTGAACATTTGAAAG 3870
3894 GTCATGATTTACAGAGAGGAAATTTAAAGCAGTAACTCTGTGCGACCGGAAAGTTT 3953
3871 GTCATGATTTACAGAGAGGAAATTCGAAACCAATCCGTGCTACGAGCGAGTAT 3930
3954 GGAACGCTGAGTCAATTTCCAGAGCAGAGCAGACCTGCGACCGGAGATGTGCAT 4013
3931 GGTCTGTATCTACCAACCTCCAGAGAGGCAAGACAGACTACCTGAGATGTCAAC 3990
4014 GCTATGGAGCATTAATCTGAGATGTGTGCAAGATAGAGACTGTACCTGACGGTCCC 4073
3991 ACACAGGCGCTTCTCCAGGCAATGTCTGCGAGAGAGAGATGTATCTTCCAGGGGCC 4050
4074 ATTTGGGCGCAATTTCTTCAAGATAGCACTTTCAACCGCTCTCTTATAGGGCGG 4133
4051 ATCTGGGCAAGATTTCAACAGAGAGGAGATTTTCAACCTCTCTCTCATAGGGTGA 4110
4134 TTTGACTCAAGAACCGGCTCTCTCAAGATCTCATCAAAACAGCGCTGTCTCGCAAT 4193
4111 TTGGACTTAAACACCTCTCTCAAGATCTCATCAAGAGACCGCGTACCTGCAAT 4170
4194 CTTCCGCGGAGTTTCAAGTCAAGATTTCTTCAATTCATCCCAATCTCAAGAG 4253
4171 CTTCCGACCACTTCAAGTGTGGCAAGTTGTCTCTTCAACAGATCTCAAGGGA 4230
4254 CA-AGTGAAGTGAATTTGAATGGAGCTGCAAGAAAGAAAGAGAGGCTGGAATCC 4312
4231 CAGGTGAGGTGAAGTTCAGTGGAGCTGCAAGAAAGAAAGAGAGGCTGGAATCC 4290
4313 CGAAGTCACTACATCAATTCATTAATGCAAAATCTGCCAAGTGTATTTACTGTGACAA 4372
4291 CGAATTCATTAATCACTTCAATCAACAGATCTGTAAATCGTGAATCCGTGATAC 4350
4373 CAATGACTTTATTAATGAGCTGCGCCCATTTGGAACCGTACTTACCTTACCGGCTGTA 4432
4351 TAAATGGGTATTTCAAGAGCTGCGCCCATTTGGAACCGTACTTACCTTACCTGTA 4410
4433 ATTAAGTGTAAATCAAAACCGGTTGATTCGTTTCAAGTTGAATTTGCTCTGCTGCT 4492
4411 ATTCGTTGTAAATCAAAACCGTTTATTCGTTGAGTGAATTTGCTGCTGCTAT 4470
4493 TCTATCTTATCGGTTTACATAGTTATAGCTTACATTAATCTGTTGCGCTTGC 4552
4471 TTTCTTCTTATTAATGATTTCAATGCTAGTGAATTAATTAACATGGGGGTTAATCAATTA 4530
4553 GATTAAGACTTACGTCATCGGGTTTACCTTATGATGAGTGGCCACTCTCTTGC 4612
4531 CTACAAAGG-----ACCCCTAGTGAAGGAGTTGGCCACTCTCTCTTGC 4574
4613 GCGTCTGCTGCTCGGTGGGGCTGCGGACCAAAAGTCCGACAGCGGAGAGCTCTGCTC 4672
4575 GCGTCTGCTGCTCACTGAGGCGGGGAGCAAAAGTCCGACAGCGGCGGCTTTCCTC 4634
4673 TGCGGCGCCACAGGAGGAGCGGAGCGGAGAGGAGGAGTGGGCA 4718
4635 GGGGCGCTCATGTAGAGGAGCGGAGCGGAGAGGAGTGGGCA 4680

```

US-08-709-609-1
? Sequence 1, Application US/08709609
? Patent No. 5858775
? GENERAL INFORMATION:
? APPLICANT: Johnson, Philip R.
? TITLE OF INVENTION: Adeno-Associated Virus Materials and
? METHODS
? NUMBER OF SEQUENCES: 3
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
? STREET: 6300 Sears Tower, 233 S. Wacker Drive
? CITY: Chicago
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/709,609
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: No. 5858775and, Greta E.
? REGISTRATION NUMBER: 35,302
? REFERENCE/DOCKET NUMBER: 31975
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (312) 474-6300
? TELEFAX: (312) 474-0448
? TELEX: 25-3856
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4680 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
US-08-709-609-1

Query Match      65.3%; Score 3082.4; DB 2; Length 4680;
Best Local Similarity 80.0%; Pred.No. 0;
Matches 3781, Conservative . 0; Mismatches 891; Indels 54; Gaps 11.

QY      1 TTGGCCATCCTCCTCGTCGCGGCTGCCTGCGTCCGTGGGGGCTCGCGACCAAGGTCCG 60
DB      1 TTGGCCATCCTCCTCTCTGCGGCTGCTGCTGACTGAAGCCCGGCGCACCAAGSTGCC 60

QY      61 AGACGGCAGAGCTCTGCTCTGCGCGGCCCCACCAGACGAGCGAGCGAGAGGAGTG 120
DB      61 CGAGCGCCCGGGGTTTGCCCGGGCGGCTCTCATGAGACGAGCGAGCGCGAGAGGAGTG 120

QY      121 GGCAACTCATTCACTAAGGGGTAATGCGAAGCGCTCCACGCTGCCGCTCAGCGCTGA 180
DB      121 GCCAACTCATCATCAAGGGGT-----TCTGAGAGGGGTGAGATCTGGA 163

QY      181 CGTAAATTACGCATAGAAGS---GAGTGGTCTGTATTAGCTGTACGTCAGAGGCTTTGG 237
DB      164 CGTGAATTACGTCATAGGGGTTTAGGAGGTCCTGTATTAGAGGTACGTAAGTG-TTTTGC 222

QY      238 GACAATTTTGGACACCAAGTGGCCATTAAAGGTATATAATGCGCAGTAGAGCAGAGGAT 297
DB      223 GACAATTTTGGCAGACCAATGTGGTCAAGCTGGGTATTAAAGCCGAGTAGCAGAGGAT 282

QY      298 CTCCATTTTGG-AACGCGAAATTGAACGAGACGACGCAATGCGGGCTTTTACGAGATCG 356
DB      283 CTCCATTTTGAAGCGGAGAGTTTGAACCGCGACCGCAATGCGGGGTTTTAACGAGATTG 342

QY      357 TGATTAAGGTGCGGAGCGACCTCGACGACGACCTGCGGCAATTTCTGATCTGTTGTGA 416
DB      343 TGATTAAGGTGCGGAGCGACCTTGAAGGCAATCTGCGGCAATTTCTGACAGCTTTGTGA 402

```

[illegible]

QY	1497	AGTCGCGCAAGCCATCTTCGCGCGGACCAAGATGCGGTGACCAAAAGGACAGTCG	1556
Db	1483	AGTCGCGCAAGCCATCTTCGAGAGACAGATGCGGTGACCAAAATGCAATCT	1542
QY	1557	CGCCGACATGACCCCAACCCCGGTGATCGTCACTCCAAACCAACAATGTGCGCGTGA	1616
Db	1543	CGGCCCATGATGACCCGACCTCCCGGATCGTCACTCCAAACCAACAATGTGCGCGTGA	1602
QY	1617	TTGACGGGAACAGCAACCACTTCGAGCAACAGACGCGGTGACGAGGACCGGATGTTCAAT	1676
Db	1603	TTGACGGGAACCTCAACGACCTTCGAAACACAGACGCGGTGACGAGGACCGGATGTTCAAT	1662
QY	1677	TTGAACTCAACCGCGGTCTGAGCACTTGGCGAAGGTGACAAAGCAGGAAGTCAAG	1736
Db	1663	TTGAACTCAACCGCGGTCTGAGCACTTGGCGAAGGTGACCAAGCAGGAAGTCAAG	1722
QY	1737	AGTTCTTCGCTGCGGCGAGATCACTGACCGAGGTGGCGCATGATCTTAAGTCAAG	1796
Db	1723	ACTTTTTCGCGGCGCAAGGATCACTGTTGAGGTGAGATGAAATTTCACTGCAAA	1782
QY	1797	AGGGTGAGCCAAACAAAGAACCAGCCCGCATGACGCGGATTAAGCAGAGCCAAAGCGG	1856
Db	1783	AGGGTGAGCCAAAGAAAGACCCGCGCCAGTGAAGCAGATTAAGTGAAGCCAAACGCG	1842
QY	1857	CCTGCCCCCTCACTCGCGAATCCATGACGTCAACGCGGAAGAGCTCCGCTGCACTTTG	1916
Db	1843	TGCGCGAATCAAGTTCCGACGACCATGACGTCAAGCGCGA--AGCTTCGATCAACTACG	1899
QY	1917	CGGACAGGTACCAAAACAAATGTTCTGTCACGCGGGGATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAAATGTTCTGTCAGTGGGATGAAATGATGCTGTTCCCT	1959
QY	1977	GCAGACATGCGAGAAATGAATCGAAATTTCAACATTTGCTTCAACGACGAGAG	2036
Db	1960	GCAGACATGCGAGAAATGAATCGAAATTTCAAAATCTGCTTCACTACGAGACAGAAAG	2019
QY	2037	ACTGTTCAAGTCTTCCCGCGGTGACAAATCTCAACCGGT--CGTCAGAAAGAGGA	2093
Db	2020	ACTGTTTGAAGCTTCTTC--CGTGCAGATCTCAACCCGTTTCTGTCTGCAAAAGG	2076
QY	2094	CGTATCGAAACTCTGTGCAATCTATCATCTGCTGGGCGGACCTCCGAGATTTGCT	2153
Db	2077	CGTATCGAAACTGTGCTACATCTCATCATCATGATGGAAGGTGCAAG--CGCTTGCA	2133
QY	2154	CGGCTCGGATTTGGTCAACGTGACCTGAGATCTGTGTTCTGAGCAATTAATGACT	2213
Db	2134	CTGCTCGGATTTGGTCAATGTGATTTGGAATGACTGCACTTTGAAACAAATTAATGATTT	2193
QY	2214	AAACAGGATATGGCTGCGCATGTTATCTTCAGATTGGCTGAGGACAACTCTCTGAG	2273
Db	2194	AAATAGGATATGGCTGCCATGTTATCTTCAGATTGGCTGAGGACAACTCTCTGGA	2253
QY	2274	GGCATTGCGAGTGTGGACTTGAACCTGAGGCCCGAAGCCCAAGCCAAACCAAGCA	2333
Db	2254	GGAAATGAACAGTGTGTGAGACTCAAACTGGCCCAACCAACCAACCCGACAGGCG	2313
QY	2334	AAGCAGGACGACGCGCGGGGTGTGGTCTCTGTGCTCAAGTACCTGAGACCTTTCAAC	2393
Db	2314	CATAGAGACGACGAGGGGTTTGTGTCTCTGGGTACAAAGTACTTGACCTTTCAAC	2373
QY	2394	GGACTCGAACAGGGGAGCCGTCACGCGCGGACGACGCGGCCCTTGAGACGACAG	2453
Db	2374	GGACTCGAACAGGGAGGCGCGGTCAACGAGGACGCGCGGCCCTTGAGACGACAA	2433
QY	2454	GGCTTACGACGACACTCAAGCGGGTGAACATCCGTATCTCGGTATTAACAGCGCGAC	2513
Db	2434	GGCTTACGACCGGCACTCGACGCGGACCAACCCGTATCTCAATTAACACAGCGCGAC	2493
QY	2514	GGCAGTTTCAGAGACGTCTGCAAAAGATAGCTCTTTTGGGGCAACCTCGGGGAGACA	2573
Db	2494	GGGAGTTTCAGAGACGCTTAAAGATATAGCTCTTTTGGGGGCAACCTCGAGACAGACA	2553
QY	2574	GTCTTTCAGGCCAGAGCGGGTTCTGAACTCTCGGTCTGCTTGAAGGAAGCGCTAAG	2633

Db	2554	GTCTTCCAGGGGAAAAAGAGGGTTCTTGAACTCTGGGCGCTGGTTGAGGAACCTGTTAAG	2613
Qy	2634	ACGGCTCTCTGGAAAGAAACGTCCGGTATGAGCACTGTGCAAGAGCCAGACTCTCTCG	2693
Db	2614	ACGGCTCCGGGAAAAAGAGGCGGTGAGCACTCTCTGGAGCCAGACTCTCTCG	2673
Qy	2694	GGCAATCGGCAAGACGAGCCAGAGCCCGCTAAAAAGAGCTCAATTTTGGTACAGCTGGC	2753
Db	2674	GGAAACCGGAAAGGGGGCCAGAGCCCTGCAAGAAAAAGATTGAATTTTGGTACAGCTGGA	2733
Qy	2754	GACTCAGATGTCAGTCCCGCATCAAACTCTCTGGAGAACTCTCCAGAACCCCGCTGCT	2813
Db	2734	GACCAGACTCAGTACTGACCCCACTCTCGAGACGCAACGAGAGCCCTCTGCT	2793
Qy	2814	GTGGAGCCTTCTCAATAGGCTTTCAGGGGATGAGCCAAATGGCAGCAATTAACGAGAGC	2873
Db	2794	CTGGGAACCTAATACGATGGCTTACAGGAGATGGCCACCAATGGCAGCAATTAACGAGAGC	2853
Qy	2874	GCCGACGAGATGGGTAAATGCTCTGAGAAATTTGGCATTTCCATATGGCTGGGCGAC	2933
Db	2854	GCCGACGAGATGGGTAAATTCCTCGGAAATTTGGCATTTCCATATGGATGGGCGAC	2913
Qy	2934	AGACTCATACCAACACACACCCCGACCTGGGGCCTTGGCCACTTAACATTAACACTCTAC	2993
Db	2914	AGACTCATACCAACACACACCCCGACCTGGGCCCTTGGCCACTTAACATTAACACTCTAC	2973
Qy	2994	AAGCAAAATCTTCAGTGGCTTCAACGGGGGCGAGCAACGACAAACAATACTTCGGCTAAGC	3053
Db	2974	AAACAAATTTCCAGCCATCA - - GAGGCTCGAAGACATCACTACTTTGGCTAAGC	3030
Qy	3054	ACCCCTGGGGGTATTTTGATTTAAACAGATTCACATGCACTTTTCAACACGTAATGG	3113
Db	3031	ACCCCTGGGGGTATTTTGATTTAAACAGATTCACATGCACTTTTCAACACGTAATGG	3090
Qy	3114	CAGGACTCATCAACAACAATTGGGATTCGGGCCCAAGAGACTCAATTTCAACTCTTC	3173
Db	3091	CAAGACTCATCAACAACAATTGGGATTCGGGCCCAAGAGACTCAATTTCAACTCTTC	3150
Qy	3174	AACATCCAAGTCAAGAGGTACGACGAAATGATGGCTCACACACTGCTTAATACCTT	3233
Db	3151	AACATTCAGATCAAGAGGTACCCGAAATGACCGGTACGACGAAATTTGCAATTAACCTT	3210
Qy	3234	ACCAAGACGGTTCAGATCTCTCGGACCTGGAGTACAGCTTCGTAAGCTCTGGGCTCT	3293
Db	3211	ACCAAGACGGTTCAGATCTCTCGGACCTGGAGTACAGCTTCGTAAGCTCTGGGCTCT	3270
Qy	3294	GCGCACGAGGCTGCTCTCTCGGTTCCTCGGAGCGTTCATGATTCGCAATACGC	3353
Db	3271	GCGCATAGAGATGCTCTCGGCTCTCGGACGACGCTTCATGATTCGCAATAGGA	3330
Qy	3354	TACCTGACGTCACAAATGAGACGCAAGCGTGGAGAGTTCATCTTTATCTGCTGAA	3413
Db	3331	TACCTGACGTCACAAAGGAGTACGAGCAAGTACGAGCTTCATTTATCTGCTGAG	3390
Qy	3414	TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTTACTTCACTACACTTTGAG	3473
Db	3391	TACTTTCTTCTTCAAGATGCTGCGTAACGGAAACAATTTACTTCACTACACTTTGAG	3450
Qy	3474	GAAATGCTTTCCACAGAGCTTACGGGCAACGACGACCTGGACCGGCTGATGAATCT	3533
Db	3451	GACGTTCTTTCCACAGAGCTTACGGGCAACGACGACCTGGACCGGCTGATGAATCT	3510
Qy	3534	CTCATTCGCAATACTGATTTTACTGTAACGAACTCAAAATCTAGTCGCGAAGTGGCCAA	3593
Db	3511	CTCATTCGCAATACTGATTTTACTGTAACGAACTCAAAATCTAGTGGGAACACACAG	3570
Qy	3594	AACAAGACTTGTCTGTTTAAAGCGTGGCTTCACAGCTGCATGTCTGTTCAAGCCCAAAAC	3653
Db	3571	CAGTCAAGAGCTTCAAGTTTCTCAAGCGGAGCGAGTGAATTCGGGACACAGTCTAGGAAC	3630
Qy	3654	TGGCTACTGGAACCTGTATTCGGCACAGCGGCTTTCTTAAACAAAAACAGACAAAC	3713

Db	3631	TGGGCTCTGGACCCTGTTAACCGCAGAGAGGAGTATCAAAAGACATCTGGGGATPAACAAC	3693
Oy	3714	AACAGCAATTTTAACTGTGACCTGTGTCTTCAAAATTTAACTCTCAATGGCCGTGAATTCATC	3773
Db	3691	AACAGTGAATCTCTGTGACTGTGAAGCTACAAAGTACCACTTCAATTTGGCAGAGACTCTCTG	3750
Oy	3774	ATCAACCTGTGCACTGTATGGCCCTCACAAAGACGACGAAGACAAGTTCTTTTCCCATG	3833
Db	3751	GTGAATCCGGGGCCCGCCATGGCAAGCCACAGAGACATGAAGAAAGTTTTCCTCAG	3810
Oy	3834	AGCGGTGCATGATTTTGTGAAAAGAGACGCGCGAGCTTCAAACTGTGATTGGACAAT	3893
Db	3811	AGCGGGGTCTCATCTCTTGGAGACAGAGCTCAGAGAAAACAAATGTGAACATTGAAAAG	3870
Oy	3894	GTCAATGATTACAGACGAGAGAGAAATTTAAAGCATTACCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCAATGATTACAGAGAGAGAAATTCGAAACAAACAAATCCGTGGCTACGAGACAT	3930
Oy	3954	GAGAACCGTGGCAGTCAATTTTCCAGACGACAGACAGACCTCGGACCGGAGATGTGCAT	4013
Db	3931	GGTTCTGTATCTACCAACCTCCAGAGAGGACAAGACAGAGACTACCGCAGATGTCAAC	3990
Oy	4014	GCTATGGAGCATTAACCTGGCATGTGTGGCAGATAGAGAGTGTACCTGCAAGGTTCC	4073
Db	3991	ACAACAAGCGTTCCTCCAGGAGATGTCTGGCAGGACAGACAGATGTGTACTTCAGGGGCC	4050
Oy	4074	ATTTGGGCGAAATTCCTCACACAGATGGACACTTACCCGTCCTCCTATGGGCGGC	4133
Db	4051	ATCTGGCAAAAGATTCCACACGAGGAGCAATTTTCAACCCTCTCCCTCATGGGTGGA	4110
Oy	4134	TTTGGACTCAAGAACCCGCTCTTCAGATCTCATCAAAAACAGCCCTGTTCTTGCAGAT	4193
Db	4111	TTCCGACTTAAACACCTCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACTTGCAGAT	4170
Oy	4194	CCTCCGGGGAGTTTTCAGCTACAAAGTTTGCTTATTCATCAACCCATACTCCACAGGA	4253
Db	4171	CTTTCGACCACTTCAGTGGCGCAAAAGTTTGCTTCTTCATCAACAAGTACTCCACGGGA	4230
Oy	4254	CA-AGTGTGTGAAATTTGAATGGAGACTGCAGAAAGAAAACAGCAAGCCTGGAAATCC	4312
Db	4231	CACGGTCAAGCTGAGATCGATGGGAGACTCGAAGAGAAAACAGCAACGCTGGAAATCC	4290
Oy	4313	CGAAGTCAGTACACATCAATTAATGCAAAATCTGCCAAGCTGATTTTACTGTGGACA	4372
Db	4291	CGAAATTCAGTACACTTCCAATCAACAAAGCTGTGTAATGTGGACTTACCGTGGATAC	4350
Oy	4373	CAATGGACTTTATCTGAGCCTCGGCCCATTTGGCACCCGTTACCTTACCGCTCCCTGTA	4432
Db	4351	TAAATGGCGTAAATTCAGAGCCTCGCCCAATTTGGCACAGATACCTGATCTGTAAATCTGTA	4410
Oy	4433	ATTACGTGTATCAATTAACCCGCTGTGATTTGCTTCACTTGAACCTTGGTCTCTGTCT	4492
Db	4411	ATTGGTTGTATCAATTAACCGTTTAATGTGGTCAATTTGAACCTTGGTCTCTGTCTGAT	4470
Oy	4493	TCTTATCTTATTCGTTTACCATGTTTATAGCTTACATTAACCTGCTTGTGCTGCTGCG	4552
Db	4471	TTCTTTCTTATCTAGTTTTCATGTGCTACGTAGATTAATAGACATGGCGGGTTAAATCA	4530
Oy	4553	GATPAAAACATTAGCTCATTCGGGTTAACCCATAGATGAGAGTTGGCCACTCCCTCTGCG	4612
Db	4531	CTACAAAGGA-----ACCCTTAGATGAGAGTTGGCCACTCCCTCTCTGCG	4574
Oy	4613	GGCGTCTGCTGCTCGGTGGGGCTTGGACCAAAAGTCCGAGACGCGAGAGCTCTGTCTC	4672
Db	4575	GGCGTCTGCTGCTCACTGAGGCGGGGCGACCAAAAGTGTGCCCGACGCGCGGGCTTTGGCC	4634
Oy	4673	TGCGGGCCCCACGAGCGAGAGCGGCGCACAGAGGAGTGGGCAA	4718
Db	4635	GGGCGGGCTCAAGTGGCCAGCGGCGCGCAAGAGAGAGTGGCCAA	4680

RESULT 5
PCT-US95-07178-1

```

: Sequence 1 Application PC/TUS9507178
: GENERAL INFORMATION:
: APPLICANT: Johnson, Philip R.
: TITLE OF INVENTION: Adeno-Associated Virus Materials and
: TITLE OF INVENTION: Methods
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
: STREET: 6300 Sears Tower, 233 S. Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/07178
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Noland, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 31975
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 474-6300
: TELEFAX: (312) 474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4680 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: PCT-US95-07178-1

Query Match      65.3%; Score 3082.4; DB 5; Length 4680;
Best Local Similarity 80.0%; Pred. No. 0;
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11

QY      1 TTGCCCACTCCCTCTCTGCGGCGCTCGCTCGGTGGGGCGCTGCGACCAAGGTCGCC 60
DB      1 TTGGGCACTCCCTCTCTGCGGCGCTCGCTCGCTCACTGAGGCGGGCGACCAAGGTCGCC 60

QY      61 AGACGGCAGAGACTCTGCTCTGCGCGGCCCGCCAGCGACGAGCGCGCGAGAGAGGAGTG 120
DB      61 CGAGCGCCCGGCGCTTGGCCCGGCGCGCTCAAGTGAGCGAGCGAGCGCGCGAGAGGAGTG 120

QY      121 GGCAACTCCATCACAAGGGGATATTCGCGAAGCGCGCTCCACAGCTGCGCGATCGAGCTGA 180
DB      121 GCCAATCTCATCACTAGGGGAT-----TCTCGAGGGGTGAGTCTGTA 163

QY      181 CGTAATTACCTCATAGG---GAGTGGTCCGTATTAGCTGTCACTGAGTGTCTTTC 237
DB      164 CGTGATTACGTCAATAGGTTTAGGAGGTCTGTATTAGAGTCACTGAGTG-TTTTGC 222

QY      238 GACATTTTGGCAGACACCACTGGGCCATTAGGGTATATATGCGCCGAGTAGCGAGCAGAT 297
DB      223 GACATTTTGGCAGACACCATGTGGTCAAGCTGGGTATTTTAAGCCCGAGTAGCAGCAGGGT 282

QY      298 CTCATTTTGG--ACCGGAAATTGGAACGAGAGAGACCAATCGCGGCTTTCAAGGATCG 356
DB      283 CTCATTTTGAAGCGGGAGGTTTGAACGCGAGCGCGCCATCCCGGGTTTTCACGATTG 342

QY      357 TGATCAAGGTCCGAGCGACCTGAGCAGACGACTGCGCGGCAATTTCGACTCGTTTGGA 416
DB      343 TGATTTAAGTCCCGAGCGACCTTGAACGGGACATCTGCCCGGCAATTTCTGACAGCTTTTGGA 402

QY      417 GCTGGGTGGCGGAGAGGATGGAGAGCTGCCCGGATTTCTGACATGAGATCTGAATCTGA 476

```

Db	403	ACTGGGTGGCCGAGAAAGAAATGGAGAGTTGGCCGCCAGATTCTGACATGATCTGAAATCTGA	462
OY	477	TTTGACAGGACACCCCTGACCGTGGCCGAGAAAGCTGACAGCGGACCTTCTGTGTCAAATGGC	536
Db	463	TTTGACAGGACACCCCTGACCGTGGCCGAGAAAGCTGACAGCGGACCTTCTGTGTCAAATGGC	522
OY	537	GCCGCGTGAATGAAGGCCCGCCGAGAGGCGCTCTTCTTTGTTCAGTTTGAGAAAGGCGAGTCT	596
Db	523	GCCGCGTGAATGAAGGCCCGCCGAGAGGCGCTCTTCTTTGTTCAGTTTGAGAAAGGCGAGAGCT	582
OY	597	ACTTCCACTCTCCATATTTCTGGTGGAGACACAGGGGGGTCAAAATCCATGGTGTGGGCGCGCT	656
Db	583	ACTTCCACTGACACGTGCTCGTGGAAACACCGGGGTAAATTCATGTTTGGAGACGTT	642
OY	657	TCCTGAGTCAGATTAGGAGCAAGCTGTGTGACAGCCATCTACCTGGCGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTGGGAAAAAAGTGAATTTACCTGGCGGATCGAGCCGACTT	702
OY	717	TGCCCAATCTGTTCCGCGGTGACCAAGACGCGTAATGGCGCGAGAGGGGGGAAACAAGTGG	776
Db	703	TGCCCAATCTGTTCCGCGGTGACCAAGAACAGAAATGGCGCGAGGCGGGAACAAGGTGG	762
OY	777	TGGACGAGTGTCAATGCCCAACTACCTCGGCCCAACAATCAGCGCGAGCTGAGTGGG	836
Db	763	TGGATGAGTGTCAATCCCAATTACTTGTCTCCCAAAACCAAGCTTGAGCTTCAGTGGG	822
OY	837	CGTGACATACATGAGAGGAGTATATATAGCGCGCTGTTTGAACCTGGCCGAGCGCAACGGC	896
Db	823	CGTGACATATATGGAACAGTATTTAAGCGCGCTGTTGAATCTCAGCGAGCGTAAACGGT	882
OY	897	TCGTGGCGCAGCACTTGACCCACGTCAGCCAGACCCAGAGATGAGAACAGAGAAATCTGA	956
Db	883	TGTGGTCCGACATCTGACGACGCGTGCAGACCGACAGACAGAAACAAAGATACGA	942
OY	957	AACCCAAATTCGAGCGCGCTGATCCGGGTCAAAAACTCCGCGCGCTACATGAGAGCTGG	1016
Db	943	ATCCCAATTCGATGTGCGCGGTGATCAGATCAAAATCTTCAGACGAGTACATGAGAGCTGG	1002
OY	1017	TCGGGTGCTGGTGGACCGGGGCATCACTCCGAGAACAGTGTATCCAGAGAGACAGG	1076
Db	1003	TCGGGTGCTGCTGGAGCAAGGGGATTACTGGAGAAACAATGTGATCCAGAGAGACAGG	1062
OY	1077	CCTGTACATCTCCTTCAACGCGCTTCCAACTGGCGGTCCCAAGATCAAGCGCGCTCTGG	1136
Db	1063	CCTCATACATCTCCTTCAATGCGGCTCCAACTCGCGGTCCCAATCAAGCGCTCCTTGG	1122
OY	1137	ACAAATGCGCGAAGATCAATGGCGGTGACCAATCCGCGGCCGACATCCTGGTAAAGCCCGG	1196
Db	1123	ACAAATGCGGAAAGATTAAGCTGATCTAAACCGCCGCCACATCACTGTGTGGACAGC	1182
OY	1197	CTCCGCCCGCGGACATTAAACCAACCCGATCTACCGCATCTTGAGCTGAAACGGCTACG	1256
Db	1183	AGCCCGTGAAGGACATTTCCAGCAATCGGATTTTAAATTTTGGAACTTAAACCGGTACG	1242
OY	1257	AACTGTCTACGCGCGCTCCGCTCTTCTCGCGCTGGGCCCAAGAAAGTTGGGAAAGCGCA	1316
Db	1243	ATCCCAATATAGCGGCTTCCGCTCTTCTGGAGTGGGCCACGAAAAAGTTCCGGCAAGAGA	1302
OY	1317	ACACCATTTGCTGTTTGGGCGCGGACACACAGGGGAAACCAACAATGCGCGGAACGCAATG	1376
Db	1303	ACACCATTTGCTGTTTGGGCGCTGACATACCGGGAAACCAACAATGCGCGGAAGCCATAG	1362
OY	1377	CCCAAGCGGTGCTTCTACAGGCTGCTCAACTGGACCAATGAGAACTTTCCCTTCATG	1436
Db	1363	CCCAACATGTGCTTCTACCGGTGCTTAACTGGAACCAATGAGAACTTTCCCTTCAAG	1422
OY	1437	ATTGCGTGCACAAAGTGTGATCTGTGTGGAGAGAGGGGCAAGATGAAGGCCAAGGTCGTG	1496
Db	1423	ACTGTGTGCACAAAGTGTGATCTGTGTGGAGAGAGGGGAAAGATGACCGCCAAGGTCGTG	1482
OY	1497	AGTCCGCCAAGGCATTTCTCGCGCGCAGCAAGTGTGCGGTGAGCCAAATAGTGCAGATGCT	1556
Db	1483	AGTCCGCCAAGGCATTTCTCGAGAGAACCAAGTGTGCGGTGAGCCCAAGAAATGCAAGTCT	1542

QY	1557	CCGCCAGATGAGACCCCAACCCCGGTGATCGTCACTCACTCCAAACCAAAATGTCGGCCGTGA	1616
Db	1543	CGGCCAGATGAGACCCCACTCCCGTGAATCGTCACTTCAACCAACCAATGTCGGCCGTGA	1602
QY	1617	TTGACGGGAACAGACCACTTCAGAGACAGACCGCTTGACAGACCAGATGTTCAAT	1676
Db	1603	TTGACGGGAATCAACGACCTTCGAAACACGACGACCGGTTGCAAGACCGATGTTCAAT	1662
QY	1677	TTGAACTCAACCCCGCTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGAGATCAAG	1736
Db	1663	TTGAACTCAACCCCGCTCTGGATCATGACTTTGGGAAGGTGACCAACACGAGAACTCAAG	1722
QY	1757	AGTCTTCGCTGGGCGAGGATCAGCGTACCGAGTGGCGCATGAGTCTCACTCAAA	1796
Db	1723	ACTTTTTCGCTGGGCGAAAGGATACGCTGGTGGAGGTGGAGCATGAATTTCACTCAAA	1782
QY	1797	AGGGTGAAGCCCAAAAGACCCGCCCCCGATGAACGGGATAAAGCGAGCCCAAGCGGG	1856
Db	1783	AGGGTGAAGCCCAAAAGACCCCGCCCCCGATGACGAGATATATAGAGAGCCCAAGCGGG	1842
QY	1857	CTGCCCCCTGATGGCGGATTCATATCGACGTCAGACGGGAAGAGGCTCCGGTGGACCTTTG	1916
Db	1843	TGCGGAGTCACTTGGCGAGCATCGACGTCAGACGGGAG--AGCTTCGATCAACTACG	1899
QY	1917	CCGACAGGTACCAAAACAAATGTTCTGTGTCACGGGCGATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAAATGTTCTGTGTCACGTTGGGCGATGATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATCAGAAATTTCAACTTTGCTTACCGCACGGGACGAGAG	2036
Db	1960	GCAAGACATGCGAGAGATGAATCAGAAATTTCAATATCTGTTACTACGACGACAGAAAG	2019
QY	2037	ACTGTTCAAGATGTTCCCGGCGCTGTCAGATCTCAACCGGT--CGTCAAGAAAGAGA	2093
Db	2020	ACTGTTTAAAGTGTCTTTC---CGTGCAGAAATCTCAACCGGTTCTGTCGTCAAAAGG	2076
QY	2094	CGTATCCGAAACTCTGTGCCATTTCATCATCTGCTGGGAGCGGGCTCCGAGATTGCTTCT	2153
Db	2077	CGTATCAAGAACTGTGTACACTTCATCATATCAATGAGGAAAGGTGCCAGAA--CGCTTGCA	2133
QY	2154	CGGCTCGGATCTGTGTCAACGCTGGAACCTGGAATGACTGTGTTCTGAGCAATTAATGACTT	2213
Db	2134	CTGCTCGGATCTGTGTCAATGTGGAATTTGGATGTGATCTGATCTTTGAAACATTAATGATTT	2193
QY	2214	AAACCGAGTATGCTGCCGATGATTTATCTTCAGATTGGCTCGAGGACAACTCTCTGAG	2273
Db	2194	AAATCAGGTATGGCTGCCGATGATTTATCTTCAGATTGGCTCGAGGACAACTCTCTGAA	2253
QY	2274	GGCATTCGCGAGTGTGTGGGACTTGAAACCTGTGAGACCCCGAAGCCCAAGCCCAACCAACCA	2333
Db	2254	GGATTAACAGTGTGTGAGACTAAACCTGTGCCACCAACCAAGCCCGCAGAGGGG	2313
QY	2334	AAGACGAGCGACGGCCGGGCTCTGATGCTTCTGCTCAACAAGTACCTTCGACCCCTTGAC	2393
Db	2314	CATAAGACGACGACGAGGGGCTCTGTGCTTCTGTGGTCAAGTACCTTCGACCCCTTGAC	2373
QY	2394	GAACTCGACAAAGGGGAGACCCCTGTCAAGCGGCGGACGACGGGCCCTTCGAGCATGACAG	2453
Db	2374	GAACTCGACAAAGGGGAGAGCCGCTCAAGAGGACACCGCGGGCTTCGAGCAAGACAA	2433
QY	2454	GCTTACGACGACGCTCAAGGCGGAGTCAAACTCGTACCTGCGGTTAAACGACGCGAC	2513
Db	2434	GCTTACGACGCGGACGCTCGACAGGAGACAACTCGTACCTCAAGTACACGACCGAC	2493
QY	2514	GCAGATTTCAAGACGCTCTGCAAGAAATACGCTTTTGGGAGCAACTTCGCGCGAGCA	2573
Db	2494	GCGAGTTTCAAGAGCGCGCTTAAGAAAGTACGCTTTTGGGAGCAACTTCGCGCGAGCA	2553
QY	2574	GTTCTTCCAGGCGCAAGAGCGGCTTCTGAACTCTTCGCTCTGTGTTGAGGAAGCGCTAAG	2633
Db	2554	GTTCTTCCAGGCGGAAAGAGGGTCTTGAACCTCTGCGCTGTGTTGAGGAACCTGTTAAG	2613

2634 ACCGCTCTCTGAAAGAAACGTCGCTAGAGAGCTGCGCACAGAGCCAGACTCTCTCTCG 2693
2614 ACCGCTCTCTGAAAGAAAGAGCGCGTAGAGCACTCTCTGAGAGAGCTCTCTCTCG 2673
2694 GGCATCGGCAAGACAGGCGCAGAGCGCGCTAAAGAGACTCAATTTTGTCAAGATGCG 2753
2674 GGAACCGGAAAGCGGCGCAGAGCGCTGCAAGAAAGATTTGATTTGTGACAGATGGA 2733
2754 GACTCAGAGTCAGTCCCGGATCCAGCAACTCTCGGAGAACCTTCAGAGAACCCCGCTGCT 2813
2734 GAGCGAGACTCAGTACCTGAGCCCCAGCTCTCGGAGAGCCACAGAGAGCCCTCTGCT 2793
2814 GTGGGACCTTACTCAATAGCTTCAAGCGCTGCGCAGCACTAATGAGCAGCAATTAAGAGG 2873
2794 CTGGGAACTAATAGAGTGTCTACAGAGAGTGGCGCAGCACTAATGAGCAGCAATTAAGAGG 2853
2874 GCCGACGAGTGGGTAAATGCTTCAAGAAATTTGCAATTTGCAATGCTGGGAGC 2933
2854 GCCGACGAGTGGGTAAATTTCTCGGAAATTTGGCATTTGCAATTTGCAATGAGTGGGAGC 2913
2934 AGAGTCACTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2993
2914 AGAGTCACTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2973
2994 AAGCAATCTCAGTGTCTTCAAGCGGCGCAGCAAGCAGCAGCAGCAGCAGCAGCAGCAGC 3053
2974 AAGCAATCTCAGTGTCTTCAAGCGGCGCAGCAAGCAGCAGCAGCAGCAGCAGCAGCAGC 3030
3054 ACCCGCTGGGGATTTTGAATTTCAACAGATTCACATGCGCACTTTTCAACAGTGTG 3113
3031 ACCCGCTGGGGATTTTGAATTTCAACAGATTCACATGCGCACTTTTCAACAGTGTG 3090
3114 CAGCGACTACCAACCAATTTGGGATTCGCGCGCAAGAGACTCAACTTCAACTTTC 3173
3091 CAAAGCTATCAACCAACCACTGGGATTCGCGCGCAAGAGACTCAACTTCAACTTTC 3150
3174 AAGTCAATCTCAGAGAGTCAAGCAGATGATGAGCTCAACCACTTCAACTTTC 3233
3151 AAGTCAATCTCAGAGAGTCAAGCAGATGATGAGCTCAACCACTTCAACTTTC 3210
3234 ACCGACGAGTCAAGTCTTCTGAGACTGAGAGTCCAGCTTCTGAGCTCT 3293
3211 ACCGACGAGTCAAGTCTTCTGAGACTGAGAGTCCAGCTTCTGAGCTCT 3270
3294 GCGCACAGAGGCTGCT 3353
3271 GCGCACAGAGGCTGCT 3330
3354 TACCTGAGCTCAACCAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3413
3331 TACCTGAGCTCAACCAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3390
3414 TATTTCCCTTCTCAGATGCTGAGAAACGCGCAACCACTTCACTTCACTTCACTTCACT 3473
3391 TATTTCCCTTCTCAGATGCTGAGAAACGCGCAACCACTTCACTTCACTTCACTTCACT 3450
3474 GAAGTCTTCTTCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3533
3451 GAAGTCTTCTTCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3510
3534 CTGATGAGCAATACCTGATTAACCTGAGCAAGCACTAATATGATGCTGGAGAGCCCA 3593
3511 CTGATGAGCAATACCTGATTAACCTGAGCAAGCACTAATATGATGCTGGAGAGCCCA 3570
3594 AAGCAAGACTTGTGTTTGAACCGTGTCTCAGCTGAGCAATGCTGTGAGCCCAAAAC 3653
3571 CAGTCAAGGCTTCAAGTTTCTCAGAGCCGAGAGCAGTACATTCGAGGACAGTCAAGAAC 3630
3654 TGGCTACCTGAGACCTGTTTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3713
3631 TGGCTACCTGAGACCTGTTTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3690
3714 AAGCAATTTTACCTGAGCAGTGTGCTTCAAAATATATACCTCAATGAGGCTGATCATC 3773

3691 AACAGTAATACCTGAGAGTGGAGTCAACAGTACCACTCAATGAGAGAGCTCTCTG 3750
3774 ATCAACCTTGGAGCTGATGAGCTTCAACAGAGCAGAGAGCAATTTTCCCATG 3833
3751 GTGAATCCGGGGCCCGCAGTGGCAAGCACAAGAGAGAGTGAAGAAATTTTCTCTAG 3810
3834 AGCGGTGATGATTTTGGAAAGAGAGCGCGAGCTTCAACATCTGATGAGCAAT 3893
3811 AGCGGTGATGATTTTGGAAAGAGAGCTTCAAGAGAAACCAATGTAATGAAATTTGAAG 3870
3894 GTGATGATTAACAGAGCAAGAGAAATTTAAAGCCATTAACCTTGTGAGCAACGAAAGATT 3953
3871 GTGATGATTAACAGAGCAAGAGAAATTTGAAGCAACCAATTCCTGTGAGAGAGTAT 3930
3954 GGAAGCTGAGCTCAATTTTCAAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4013
3931 GGTCTGTATCTTCAACCTTCCAGAGAGGCAACAGCAGCAGCAGCAGCAGCAGCAGC 3990
4014 GGTATGAGCAATTAACCTGAGAGTGTGAGCAAGATGAGAGAGTATACCTGAGAGGCTCC 4073
3991 AAGCAAGGCTTCTTCAAGCAGTGTGAGCAGAGAGAGATGATCTTCAAGAGGCTCC 4050
4074 ATTTGGGCAAAATTTCTTCAACAGAGATGAGCACTTCAACCGCTCTCTCTATGAGGAGC 4133
4051 ATCTGGGCAAAAGTTCACAGAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4110
4134 TTGGAATCAAGAGCCGCTCTCTCAATCTCATTAAGAAACAGCGCTTCTCTGAGAT 4193
4111 TTGGAATCAAGAGCCGCTCTCTCAATCTCATTAAGAAACAGCGGATCTTGGAAAT 4170
4194 CTTGGGAGAGTCTTCAAGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4253
4171 CTTGGAATCAAGAGCCGCTCTCTCAATCTCATTAAGAAACAGCGGATCTTGGAAAT 4230
4254 CA-AGTATGAGTGAATTTGAATGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGATCC 4312
4231 CAGGTCAGAGTGTGAGATGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGATCC 4290
4313 CGAAGTGTGAGTCAATTCATTAATGAGAAATGTGCAACGTTGATTTTCTGTGAGCA 4372
4291 CGAATTTGAGTCAATTCATTAATGAGAAATGTGCAACGTTGATTTTCTGTGAGTAC 4350
4373 CAATGAGCTTAACTGAGAGCTGCGCCCAATGAGCAACCGTATCTTAACCGGCTCCGTA 4432
4351 TAATGAGCTGTATTAAGAGCTGCGCCCAATGAGCAACCGTATCTTAACCGGCTCCGTA 4410
4433 ATTAAGTGTATTAATCAATTAACCGTGTGATTTGATTTGATTTGATTTGATTTGATTTG 4492
4411 ATTAAGTGTATTAATCAATTAACCGTGTATTTGATTTGATTTGATTTGATTTGATTTG 4470
4493 TCTTAATCTTAATGAGTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4552
4471 TCTTAATCTTAATGAGTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4530
4553 GATTAAGATTAATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4612
4531 GATTAAGATTAATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4574
4613 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4672
4575 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4634
4673 TGCAGGCCCCAGCAG 4718
4635 GCGCGGCTCTGAGTGAAG 4680

RESULT 6
US-09-770-315-3
; Sequence 3, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:

```

1  APPLICANT: Chiron Corporation
2  TITLE OF INVENTION: Recombinant AAV Packaging Systems
3  FILE REFERENCE: 20263-501
4  CURRENT APPLICATION NUMBER: US/09/770,315
5  CURRENT FILING DATE: 2001-01-26
6  PRIOR APPLICATION NUMBER: US 60/178,536
7  PRIOR FILING DATE: 2000-01-26
8  NUMBER OF SEQ ID NOS: 8
9  SOFTWARE: Fastseq for Windows Version 3.0
10 SEQ ID NO 3
11 LENGTH: 7557
12 TYPE: DNA
13 ORGANISM: Unknown
14 FEATURE:
15 OTHER INFORMATION: recombinant DNA
16 US-09-770-315-3

```

Query Match	63.4%	Score 2993.4	DB 4	Length 7557
Best Local Similarity	80.4%	Pred. No. 0		
Matches 3653; Conservative	0	Mismatches	841	Indels 45; Gaps 9

[illegible]

QY	769	CAAGTGGTGGACACAGTGTCTACATCTCCCACTACTCTCTGCCCCAAGACTCAGCCCGAGCT	828
Db	783	CAAGGTGGTGGATGGTGTCTACTCTCCCAATTACTTGTCTCCCAAAACCCAGCTGAGCT	842
QY	829	GCAGTGGGCGTGGACCTTAACATGAGAGAACTATATAAGCCCTGTTTGAACCTGGCGAGCG	888
Db	843	CCAGTGGGCGTGGACTAATATATGGAAACATATTTTAAGCGCCGTGTTGAATCTCAGCGAGCG	902
QY	889	CAATCGGCTCGTGGCGAGCACCTGACCACGTCAGCCAGAACCCAGAGCAACACAGGA	948
Db	903	TAAAGGTTGGTGGCGAGCACTCTGACCAAGTGTGAGACGCGAGGAGCAACAAAGA	962
QY	949	GAATCTGAACCCCAATTTCTGAACGCGCCGTATCCGGTCAAAAACCTCCGCGCTACAT	1008
Db	963	GAATCAAGAAATCCCAATTTCTGAATGCGCCGGTATCAGATCAAAAATTTAGCCAGTACAT	1022
QY	1009	GGAGCTGGTGGGTGGCTGTGGAACCGGGGANTCACTCCGAGAAAGCTGATTCAGGA	1068
Db	1023	GGAGCTGGTGGGTGGCTGTGGAACAAAGGGATTACCTCGAGAAAGCTGATTCAGGA	1082
QY	1069	GGACCAAGGCTTCGTACATCTCTTCAACAGCCGTTCCAACTCCGGGTCCAGATCAAGCC	1128
Db	1083	GGACCAAGGCTTCGTACATCTCTTCAAGCGGCTCCCACTCCGGGTCCCAATCAAGGC	1142
QY	1129	CGCTCTGACATGCGCGCAAGATCATGGCCCTGACCAAAATCCGCGCCGACTACTGGT	1188
Db	1143	TGCTTGGACATGGGGAAAGATTATGAGCTGACTAAACCGGCCGACTACTGGT	1202
QY	1189	AGGCCCCGCTCGGCCCGGACATTAAACCAACCGCATCTACCGCATCTGGAGCTGAA	1248
Db	1203	GGGCGAGAGCGCCGTGGAGGACATTTCCAGCAATCGGATTTATATAATTTTGGAACTAA	1262
QY	1249	CGGCTACCAACCTGCTTACGCGGCTCCGTCTTCTCGGCTGGGCCCCGAAGAGTTTCG	1308
Db	1263	CGGCTACCAATCCCAATATGCGGCTCCGTCTTCTCGGAGTGGGCGACGAAAGTTTCG	1322
QY	1309	GAAGCGCAACCATCTGGCTGTGTTGGGCGGACCAACGCGCAACCAACATCGCGGA	1366
Db	1323	CAAGGGAAACACATCTGGCTGTGTTGGGCTCTGCAACTACCGGAAAGACCAACATCGCGGA	1382
QY	1369	AGCCATCGCCACGCGCGTGGCCCTTCAACGGCTGGGTCAACTGGAACCAATGGAATTTCC	1428
Db	1383	GGCATTAGCCCACTGTGTCCCTTCTACGGGTGCTTAACCTGGAACCAATGGAATTTCC	1442
QY	1429	CTTCATGATTTGCTGCAACAGATGATGTGATCTGTGGAGGAGGGCAAGATGACGGCCAA	1488
Db	1443	CTTCACACACTGTGTGCAACAGATGATGTGATCTGTGGAGGAGGAGGAAGATGACGGCCAA	1502
QY	1489	GGTGTGAGATCGGCGCAAGGCAATTTCTCGGCGGAGCAAGGTGGCGGTGACCAAAAGTG	1548
Db	1503	GGTGTGAGATCGGCGCAAGGCAATTTCTCGGAGGAAGCAAGGTGGCGGTGACCAAAAGTG	1562
QY	1549	CAAGTCGCCCCAGATCGACCCCAACCCCGTATCGTCACTCTCAACCAACATGTG	1608
Db	1563	CAAGTCCTCGGCCAGATAGACCCGACCTCCGTATGTCACTCTCAACCAACATGTG	1622
QY	1609	CGCGGTGATTAACGGGAAACAGACCACTTCTGAGACCAAGACCGCTTTCAGACCGGAT	1668
Db	1623	CGCGGTGATTAACGGGAACTCAAGACCTTCTGAAACACAGACCGCTTTCAGACCGGAT	1682
QY	1669	GTTCAAAATTTGAACCTCACCGCGCTGTGGATCATGACTTTGGGAAGGTCAACCAAGCAGA	1728
Db	1683	GTTCAAAATTTGAACCTCACCGCGCTGTGGATCATGACTTTGGGAAGGTCAACCAAGCAGA	1742
QY	1729	AGTCAAAAGTTCTTCGCTGGGCGCAGGATCACGTGACCGAGGTGGCGCATGATTTCTA	1788
Db	1743	AGTCAAAAGTTCTTCGCGTGGGCAAAAGATCAAGTGGTGAAGGTGAGCATGATTTCTA	1802
QY	1789	CGTCAAAAGGTGGAGCAACAAAGACCGCGCCCGGATGAGCGCGATTAAGACGAGCC	1848
Db	1803	CGTCAAAAGGTGGAGCAAGAAAGACCGCGCCCGGATGAGCGCGATTAAGTGAAGCC	1862
QY	1849	CAAGCGGCGCTGCGCTCAGTGTGGGATTCATCAAGTCAAGCGCGAAGAGGCTTCGGT	1908

QY 4066 AGGCTCCATTTGGGCGCAAAATTCCTCAACAGATGGAACCTTCAACCCGCTCTCTTA 4125
DB 4071 AGGGGCGCATCTGGGCAAAAGATTCACACGAGCGACATTTTACCCCTCTCCCTCA 4130
QY 4126 TGGGCGGCTTGGACTCAAGAACCCGCTCTCTCAGATCTCTCATCAAAAACAAGCTGTTC 4185
DB 4131 TGGGATGATTCGACTTAACACCCCTCTCCACAGATTTCTCATCAAGAACACCCGGTAC 4190
QY 4186 CTGCGAATCTCCGCGGAGTTTTCAGCTACAAAGTTTGTCTTCAATCATCAACCAATACT 4245
DB 4191 CTGCGAATCTCTTCAACCACTTCAGTGGCGCAAAAGTTTGTCTTCAATCATCAACAGTACT 4250
QY 4246 CCACAGGCAAGTGAAGTGTGAATAATGAAATGAGAGCTGACAGAAAGAAAGCAAGAGGCT 4305
DB 4251 CCACGAGCAGGTCAGCTGAGTGAATCGATGGAGCTGACAGAAAGAAAGCAAGAGGCT 4310
QY 4306 GGAATCCGAGAGTCAAGTACATCAATTAATGCAAAATCTGCCAAGCTTGAATTTTACTG 4365
DB 4311 GGAATCCGAGAAATTCAGTACATTCACATCAAGTCTTAAATGTGACTTTTACTG 4370
QY 4366 TGAACAAATGACCTTAACTGAGCTCGCCCATTTGGACACCCGTTACTTACCCGTC 4425
DB 4371 TGAACAAATGAGGCTGATTAACAGAGCTCGCCCATTTGGACACCAATACCTGACTGTA 4430
QY 4426 CCCTGTAATTAAGTGAATCAATTAACCGGTTGATTTGTTCAAGTTGAACCTTGTGCTC 4485
DB 4431 ATCTGTAATTCCTGTTATCAATTAACCGTTTAATTCGTTTCAAGTTGAACCTTGTGCTC 4490
QY 4486 CTGCTCTTCTTATCTTATC 4504
DB 4491 TGGCATTTCTTCTTATC 4509

RESULT 7
US-09-532-594B-1
Sequence 1, Application US/09532594B
Patent No. 6468524
GENERAL INFORMATION:
APPLICANT: Chorini, John A.
APPLICANT: Kotin, Robert M.
APPLICANT: Safier, Brian
APPLICANT: Davidson, Beverly
TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
FILE REFERENCE: 14014.0252U2
CURRENT APPLICATION NUMBER: US/09/532,594B
CURRENT FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 4767
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
OTHER INFORMATION: synthetic construct
NAME/KEY: misc feature
LOCATION: 3009
OTHER INFORMATION: n = g, a, c or t (u)
NAME/KEY: misc feature
OTHER INFORMATION: AAV4 genome
US-09-532-594B-1

Query Match 55.0%; Score 2596; DB 4; Length 4767;
Best Local Similarity 74.5%; Pred. No. 0;
Matches 3582; Conservative 0; Mismatches 1096; Indels 129; Gaps 20;

QY 1 TTGCGCACTCCCTCTCTGCGGCTCGCTCGGCTGGGCGCTCGGACCAAGATCGGC 60
DB 1 TTGCGCACTCCCTCTCTGCGGCTCGCTCGGCTGGGCGCTCGGACCAAGATCGGC 60
QY 61 AGACGCGAGAGCTGTGCTCTCCGCGCCACCGAGCGAGCGCGCGAGAGGAGTG 120
DB 1 AGACGCGAGAGCTGTGCTCTCCGCGCCACCGAGCGAGCGCGCGAGAGGAGTG 120

DB 61 AGACGCGAGCTGTGCGCGGACAGGCGGAGTGAAGAGGAGCGGCATAGAGGAGTG 120
QY 121 GCGAATCCATCTAGTGGG-----GTAATCGGGAAGCGGCTCCAC 161
DB 121 GCGAATCCATCTAGTGGG-----GTAATCGGGAAGCGGCTCCAC 180
QY 162 GTCGCGCGTCAAGCGTGAAGTAAATTAAGTCAATAGGAGAGTGTCTGTGTA-----T 213
DB 181 GTCGCGTATTAAGTCAAGTGAAGTGTCTATTTGCGGAGGAGTGAAGGAGCATACC 240
QY 214 TACCTGTACGTGA-----GTGCTTTTCCAGATTTTCCAGACCACTGTGCCAT 263
DB 241 TACCTGTACGTGA-----GTGCTTTTCCAGATTTTCCAGACCACTGTGCCAT 300
QY 264 TTAAGGATATTAAGGCGAGTGAAGGAGC-AGGATCTCCATTTTGAACCGGAAATTTGA 322
DB 301 G-AGGATATTAAGGCGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 359
QY 323 CGAGCAGCAGCAGTGGCGGCTTCTACAGATCTGATCAAGGTTGCGAGCAGCTTGAC 382
DB 360 CGAGCAGCAGCAGTGGCGGCTTCTACAGATCTGATCAAGGTTGCGAGCAGCTTGAC 419
QY 383 GAGCAGCTGCGGCGATTTCTGACTGCTTTGAGCTGGGTGCGGAGAAAGATGGAG 442
DB 420 GAGCAGCTGCGGCGATTTCTGACTGCTTTGAGCTGGGTGCGGAGAAAGATGGAG 479
QY 443 CTGCGCGGATTTCTGACATGATCTGAATCTGATTTGAGAGGAGCAGCCTGAGCGGCG 502
DB 480 CTGCGCGGATTTCTGACATGATCTGAATCTGATTTGAGAGGAGCAGCCTGAGCGGCG 539
QY 503 GAGAGCTGCGGCGGAGCTTCTGATCAATGAGCGCGCGGAGTGAAGGCGCGGAGGCG 562
DB 540 GAGAGCTGCGGCGGAGCTTCTGATCAATGAGCGCGCGGAGTGAAGGCGCGGAGGCG 599
QY 563 CTCTCTTTTGTAGTTGAGAAAGGCGAGCTCTTCACTTCCATATTTCTGTGAG 622
DB 600 CTCTCTTTTGTAGTTGAGAAAGGCGAGCTCTTCACTTCCATATTTCTGTGAG 659
QY 623 ACCAGCGGAGTCAATTCATGATGAGGCGGCTCTCTGATGAGATTAAGGAGCAAGCTG 682
DB 660 ACCAGCGGAGTCAATTCATGATGAGGCGGCTCTCTGATGAGATTAAGGAGCAAGCTG 719
QY 683 GTGCAAGCATCTACCGCGGAGTGAAGCGGAGCCTGCGCAATGTTGCGGATGACCAAG 742
DB 720 GTGCAAGCATCTACCGCGGAGTGAAGCGGAGCCTGCGCAATGTTGCGGATGACCAAG 779
QY 743 ACCGCTATGCGCGCGGAGGAGGAGCAAGGTTGAGAGGAGTCTTCCCACTAC 802
DB 780 ACCGCTATGCGCGCGGAGGAGGAGCAAGGTTGAGAGGAGTCTTCCCACTAC 839
QY 803 CTCTGCGCAAGCTAGCGCGGAGTGAAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAG 862
DB 840 CTCTGCGCAAGCTAGCGCGGAGTGAAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAG 899
QY 863 AGCGCTGTTTGAATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922
DB 900 AGCGCTGTTTGAATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 959
QY 923 AGCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982
DB 960 TCGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1019
QY 983 CGGTCAGAAATCTCCGCGGCTTACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1042
DB 1020 AGGTCAGAAATCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1079
QY 1043 ACCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1102
DB 1080 ACCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1139
QY 1103 TCGAATCTGCGGCTTCCAGATCAAGGCGCTTGAAGCAATGCGGAGGAGGAGGAGGAGG 1162
DB 1140 TCGAATCTGCGGCTTCCAGATCAAGGCGCTTGAAGCAATGCGGAGGAGGAGGAGGAGG 1199

QY 1163 ACCAAATCCGGCCCGACTACTGCTAGGCCCCGCTCCGCCGGAATTAAACCAAC 1222
DB 1200 ACAAAGCGGCTCCGACTACTGCTAGGCCCCGGAACCCGCCGGAATTTCCAGAAC 1259
QY 1223 CGCATCTACCGCATCTCTGAGCTGAACCGGCTACGAACTTGCTACCGCGGCTCCGCTTTT 1282
DB 1260 CGCATCTACCGCATCTCTGAGCTGAACCGGCTACGAACTTGCTACCGCGGCTCCGCTTTT 1319
QY 1283 CTGGGCTGGGCGGCAAAAGGTTCCGGGAAGCGCAACCATCTGGGCTTTTGGGCGGCGC 1342
DB 1320 CTGGGCTGGGCGGCAAAAGGTTCCGGGAAGCGCAACCATCTGGGCTTTTGGGCGGCGC 1379
QY 1343 ACCACGGGCAAGACCAACATCTGCGGAAGCATGCGCCACCGCTGCTCTTCAAGGCTG 1402
DB 1380 ACGACGGGTAAACCAACATCTGCGGAAGCATGCGCCACCGCTGCTCTTCAAGGCTG 1439
QY 1403 GTCACTGACCAATGAGAACTTTCCCTCAATGATTTGCTGCAAAAGATGATCTGG 1462
DB 1440 GTGAACCTGACCAATGAGAACTTTCCCTCAATGATTTGCTGCAAAAGATGATCTGG 1499
QY 1463 TGGGAGAGGGCAAGATGAGCGCCCAAGTGTGAGTCCGCAAGGCTATTCGAGCGC 1522
DB 1500 TGGGAGAGGGCAAGATGAGCGCCCAAGTGTGAGTCCGCAAGGCTATTCGAGCGCA 1559
QY 1523 AGGAAGGTGGCGGTGACCAAAAGTGAAGTGTGCGCCGACATCCACCCCGCTG 1582
DB 1560 AGGAAGGTGGCGGTGACCAAAAGTGAAGTGTGCGCCGACATCCACCCCGCTG 1619
QY 1583 ATGTCACCTTCCAAACCAACATCTGCGCGGTGATTAAGCGGAACAGACACCTTTCAG 1642
DB 1620 ATGTCACCTTCCAAACCAACATCTGCGCGGTGATTAAGCGGAACAGACACCTTTCAG 1679
QY 1643 CACGAGGCGGCTTGCAGAGCCGGAATGTTCAATTGAATCAACCGCGCTCTGAGCAT 1702
DB 1680 CACGAGGCGGCTTGCAGAGCCGGAATGTTCAATTGAATCAACCGCGCTCTGAGCAT 1739
QY 1703 GACTTGGCAAGGTGACCAAGAGCAAGAGTCTTCCGCTGGGCGGCAAGATCAC 1762
DB 1740 GACTTGGCAAGGTGACCAAGAGCAAGAGTCTTCCGCTGGGCGGCAAGATCAC 1799
QY 1763 GTGACCGAGGTGGCGCATGAGTCTTCACTGACAGAAAGGTGAGCCCAAAAGACCGGCC 1822
DB 1800 GTGACCGAGGTGACCTCAAGATTTTACGTAGAAAGGTGAGCTAGAAAGAGCCCGGCC 1859
QY 1823 CCGGATGACCGGATTAAGAGCGAGCCCAAGCGGCTGCGCTCAAGTCCGAGTCCATCG 1882
DB 1860 CCGGATGACCGGATTAAGAGCGAGCCCAAGCGGCTGCGCTCAAGTCCGAGTCCATCG 1919
QY 1883 AGGTGACGCGGAGAGGCTCCGGTGGACTTTGCGGACAGGTACCAAAACAATGTTCT 1942
DB 1920 AGGTGACGCGGAGAGGCTCCGGTGGACTTTGCGGACAGGTACCAAAACAATGTTCT 1976
QY 1943 CGTACGCGGCGCATGCTTCAAGTCTGTTTCCCTGCAAGACATGCGAGAAATGAATCAG 2002
DB 1977 CGTACGCGGCGCATGCTTCAAGTCTGTTTCCCTGCGGCAATGCGAGAAATGAATCAG 2036
QY 2003 AATTTCACATTTGCTTCAAGCAGCAGGGAAGAGATCTTTCAGATGCTTCCCGGCGTG 2062
DB 2037 AATTTCACATTTGCTTCAAGCAGCAGGGAAGAGATCTTTCAGATGCTTCCCGGCGTG 2093
QY 2063 TCAGAAATCTCAAC-----CGGTGTCAGAAAGAGAGCGTATGGGAAATCTGTGCAATT 2116
DB 2094 TCAGAAATCTCAACCGGTCGTCTGTCTGTCAGAAAGAGAGCGTATGAAATCTGTGCAATT 2153
QY 2117 CATCATCTGCTGGGCGGAGCTCCCGAATGCTTGTCTCGGCTGCGATCTGTCAACGTG 2176
DB 2154 CATCATCTGCTGGGCGGAGCGCGCGAGAGTGGCTGCTCGGCTGCGAATGCGCAATGTG 2213
QY 2177 GACTTGATGACTGTGTTTCTGAGCAATAATGACTTAAACCAAGATATGCTGCGCATG 2236
DB 2214 GACTTGATGACTGTGACTGATGAAACAAATAATGACTTAAACCAAGATATGACT---GACGG 2270

QY 2237 TTATCTTCAGATTTGGCTGAGAGCAACCTCTCTGAGGCGATTCGGAAGTGTGGACATT 2296
DB 2271 TTATCTTCAGATTTGGCTGAGAGCAACCTCTCTGAGGCGATTCGGAAGTGTGGACATT 2330
QY 2297 GAAACTTGAAGCCCGGAGAGCCCAAGGCAACCAAGCAAAAGCAGAGAGAGCGCGGCTCT 2356
DB 2331 GAAACTTGAAGCCCGGAGAGCCCAAGGCAACCAAGCAAAAGCAGAGAGAGCGCGGCTCT 2390
QY 2357 GGTGCTTCCGGCTACAGACTGAGGAGCCCTTCAACGAGCTGCAAGAGGAGGAGCCGT 2416
DB 2391 TGTGCTTCCGGCTACAGACTGAGGAGCCCTTCAACGAGCTGCAAGAGGAGGAGCCGT 2450
QY 2417 CAACGCGGCGAGACGACGCGCCCTCGAGACGACCAAGGCGCTACAGACAGCTCAAGC 2476
DB 2451 CAACGCGGCGAGAGCGGAGAGCCCTCGAGACGACCAAGGCGCTACAGACAGCTCAAGC 2510
QY 2477 GGGTGAACAATCCGTAACCTGCGATTAACAAGCGCGAGCTTTCAGAGAGCTTGA 2536
DB 2511 CGGTGAACAACCCCTACCTCAAGTACCAACGCGCGAGGTTTCAGAGAGCTTGA 2570
QY 2537 AGAAGTACGCTTTTGGGGGCAACCTCGGGGAGAGCTTTCAGAGCAAGAGGAGG 2596
DB 2571 GGGCGACACACGTTTGGGGGCAACCTCGGGGAGAGCTTTCAGAGCAAGAGGAGG 2630
QY 2597 TCTGAACTCTCTGCTGTGTTGAGAGAGCGCTTAAGACGCTCTTGAAGAAAGCTCC 2656
DB 2631 TCTGAACTCTCTGCTGTGTTGAGAGAGCGGTTGAGAGAGGCTCTTGAAGAAAGAGCTC 2690
QY 2657 GGTGAGCAGTGGCCCAAGAGAGCAGATCTCTCTCGGCGATGCGCAAGAGCGGCA 2716
DB 2691 GTTGATTTGATCCCGCCACAGCGCCGATCTCTCAAGGATATGCGCAAAAGAGCA 2750
QY 2717 GCGGCTTAAAGAGCTCAATTTGTCAGACTGCGCGCTCAAGTCTGCTCCGATCC 2776
DB 2751 GCGGCTTAAAGAGCTCAATTTGTCAGACTGCGCGCTCAAGTCTGCTCCGATCC 2801
QY 2777 ACNACCTCTCGAGAACTCTCAGCAACCCCGCTGCTGTGAGCACTTCAATGAGCTTC 2836
DB 2802 ACCCGCTGAGGATCAACTTCCGAGACCAATGTCATGACAG-----TGAATGAGTGC 2855
QY 2837 AGGCGGTGGGCAACCAATGCGAGCAATTAACAGAGCGCGGACGAGTGGGTATGCTTC 2896
DB 2856 AGGCGGTGGGCAACCAATGCGAGCAATTAACAGAGCGCGGACGAGTGGGTATGCTTC 2915
QY 2897 AGGAATTTGCGATTTGATTTCAACATGCGGCGGCAAGATATCAACCAAGCAACCG 2956
DB 2916 GGGTGAATTTGCGATTTGATTTCAACATGCGGCGGCAAGATATCAACCAAGCAACCG 2975
QY 2957 CACTTGGGCTTGGCCACCTACCAATTAACCACTCTCAAGCAAAATCTCCAGTGCCTCAAC 3016
DB 2976 CACTTGGGCTTGGCCACCTACCAATTAACCACTCTCAAGCAAAATCTCCAGTGCCTCAAC 3024
QY 3017 GGGGAGCAGCAACGACCAACCACTTCTGAGTCAAGACCCCTGAGGATTTGATTT 3076
DB 3025 -GAGAGCTGAGTCAACCAACCACTTCTGAGTCAAGACCCCTGAGGATTTGATTT 3083
QY 3077 CAACCAATTTCACTGCACTTTTCAACAGTGAATGAGGAGCACTATCAACCAATTTG 3136
DB 3084 CAACCAATTTCACTGCACTTTTCAACAGTGAATGAGGAGCACTATCAACCAATTTG 3143
QY 3137 GGGATTCGCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCAAGTCAAGAGGCTAC 3196
DB 3144 GGGATTCGCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCAAGTCAAGAGGCTAC 3203
QY 3197 GACGAATGATGGGCTCAACCAATGCTTAATACTTCAACAGACGCTTCAAGTCTTCTC 3256
DB 3204 GACGTGAAGCGGAGAGCAACGCTGCTAATACTTCAACAGACGCTTCAAGTCTTCTC 3263
QY 3257 GACGTGAAGTCAACACTTCCGTAAGTCTCTGCGCTTGGGCAACCAAGGCTGCTCCCTCC 3316
DB 3264 GACGTGCTGAAGTCAACACTTCCGTAAGTCTCTGCGCTTGGGCAACCAAGGCTGCTCC 3323
QY 3317 GTTCCGCGGAGAGTGTATGATTTCCGAATACGCTAC-----CTGACGCTCAA 3367

Db	3324	TTTTCCCAACGACGCTTTATGTCGCCCACTACGCGTACCTGTGACCTGGAC	3383
Qy	3368	CAATGGCAGCCAGACCGTGGGACGTTCACTCTTTACTGCTCGAATATTTTCCCTTCA	3427
Db	3384	CACCTCGCAGCAACAGACTGACAGAAATGCTTCACTGCTCGGAGTCTTTCCTCGCA	3443
Qy	3428	GATGCTGGAACGGGCACAACCTTACTCTCAGCTACACCTTTGAGGAAGTGCCTTCCA	3487
Db	3444	GATGCTGGGACTGGCAACAATTGAAATTCGACGTTTGTGAAGGTGCCTTTCCA	3503
Qy	3488	CAGCAGCTACCGCGCACAGCCGAGCGTGGACCGGCTGTATGATCTCTCATCGACATA	3547
Db	3504	CTCGATGTACGGGACACAGCCAGAGCTTGGACCGGCTGTATGAACCTCTCATCGACAGTA	3563
Qy	3548	CCTGTATTACTGTAAACAGAACTCAAAATCACTGTCGGGAA--AGTGCCTCAAAACAGAACTT	3604
Db	3564	CCTGGGGGACCTGCATTCAGACACACACCGGAAACAACCTGTATGTCGGGAACTGCCACAC	3623
Qy	3605	GCTGTTTATCCGTTGGTCTCCACAGCTGGGACATGCTGTAGCCCAAAAAGTGGTCACTGG	3664
Db	3624	CAACTTTACCAAGCTGCGGCTTACCAACTTTTCAAATTTAAAAAGACTGGCTCCCGG	3683
Qy	3665	ACCCTGTTATCGCGACGACGCGCTTTCPAAAACAAAACAGACAAACAAACAGCAATTT	3724
Db	3684	GCTTTCATCAAGCAGCAGGCGCTTCTCAAAGCTGCCAATCAAACTCAAGATCTCTGC	3743
Qy	3725	TACCTGGACTGTGCTGTTCAAAATATAAC-----CTCATGGGCGGTGATC	3769
Db	3744	CACGGGGTCAAGCACTCTCATCAATACGAGACGACACACTTGGACGGAAATGGAG	3803
Qy	3770	CATCATCAACCTTGCGACTGTATGGCTCAACAAGACGACGAAAGCAAGTTCTTTCC	3829
Db	3804	TGCCCTGACCCCGGACCTCCCAATGGCCACGCGCTGGACCTGGGACAGCAAGTTACG---	3866
Qy	3830	CATGACGGGTGTCAATTTTTTGGAAAAAGAGCGCCGGAGCTTCAACACTGACTTTGGA	3889
Db	3861	CAACGACGAGCTCACTTTTGGCGGGCTTAAACAGAACGGCAACAGCGCACCGTACCCGG	3920
Qy	3890	CAATGTCAATGATTACAGACGAAAGAAATTTAAAGCCACTAACCTGTGGCCACCGAAG	3949
Db	3921	GACTCTGATCTTCACTCTGTAGAGAGACTGGCACCTCAACCGCACCGATACGAGAT	3980
Qy	3950	ATTGGGACCGTGGCACTCAATTTCCAGACGACGACACAGACCTTGGCACCGGAGATGT	4009
Db	3981	GTGGGGCAACTCACTCTGGCGGTGACCAAGCAACGCAACTGTCCGACCGTGGACAGAT	4040
Qy	4010	GCATGCTATGGGAGCACTTACCTGGCATGTGTGGCAATATGAGACGTTACCTGGACGG	4069
Db	4041	GACACCTTGGGAGCGTGCCTGGAGATGCTGTGGAAAAACAGACATTTTACTACACGGG	4100
Qy	4070	TCCCATTTTGGGCAAAATTTCTCCACACAGATGACACTTTCACCCGTCCTCTTATGGG	4129
Db	4101	TCCCATTTTGGGCAAGATTTCTCATACCGATGACACTTTCACCCCTCACCGCTATTTGG	4166
Qy	4130	CGGCTTTGGAATCAAGAAACCGGCTCTCTCAATCTCTCATCAAAAACAGGCTGTTCCTGC	4189
Db	4161	TGGGTTTGGGCTGAAACACCGGCTCTCTCAATTTTATCAAGAACACCCCGTACTCGC	4220
Qy	4190	GAATCTCCGGCGGAGTTTTCAGCTACAAAGTTTTCCTTCAATTCATCAACCAATCTCCAC	4249
Db	4221	GAATCTGCAACGACCTTCAGCTCACTCCGGTAAACCTTTCAATTACTCAGTACAGAC	4280
Qy	4250	AGGACAGTGAAGTGTGAAATTTGAATGGGAGCTGCAGAAAGAAAAACAGCAAGCGTGGAA	4309
Db	4281	TGGCAGAGGTGTGGGCGAGATTGATCTGGGAGATTCAGAAAGGACGGGTCAAAACGCTGGAA	4340
Qy	4310	TCCGCAAGTGAATCACATCAATTAATGCAAAATCTGCAACGTTGATTTTACTGTGGA	4369
Db	4341	CCCGAGGTCAAGTTTACTCTCAACTACGACAGCAAAATCTCTGTGTGGGCTCCCGA	4400
Qy	4370	CAACATGGAATTTTACTGAGCCTGCGCCCAATTGGCAACCGGTACCTTACCCTGCTCT	4429

D	b		4401	TGCGGCCTGGGAATATACCTAGACTAGCCGTATGCCATGCCAGCCGGTACTCAACCACAACCT	4468
Oy			4430	GTAATTTCGTTTATCATATAAACCCGGTAGTTTGTTTCACTGAATTTGGTCTCC-TG	4488
D	b		4461	GTAATATCCTGTTAATCAATAAAACCGGTTTATTCGTTCAGTTGTAATTTGGTCTCGTG	4520
Oy			4489	TCCTCTTATCTTATTC-CGTTACCATGGTATAGCTTACACATTAATCTG-----CTYGG	4541
D	b		4521	TCCTCTTATCTTATCTTCGTCTTTCCATGGCTACAGTCCGATAGAAGCACACGGCCCTGCCGGCC	4580
Oy			4542	TTGGCGTTTCGCGATAAAAGACTTACG-----TCATCGGGTTACCCCTAGTAGTG	4591
D	b		4581	TTGGCGTTTCGCGGTTTTCAACTGCGCGTTAATCATGTAACCTTCGGCAAACAGATGATGG	4640
Oy			4592	AGTTGGCCACCTCCCCTCTCGCGGCGCTGCTGCTCGGTGGGGGCGCTGGGACCAAAGGTCC	4651
D	b		4641	AGTTGGCCAACTTAGCTATGCGGCTGCTCACTCACTCGGCTTGAGAGACCAAGGTCT	4700
Oy			4652	GCAGACGCGACAGAGCTCTGCTCTGCGCGCCCCACCGACGAGCGAGCGCGAGAGGGAG	4711
D	b		4701	CCAATACCTCGCGGCTCTGCGCGGACGGCCGAGTGAAGTAGCGCAGCGCGCATAGAGGGAG	4760
Oy			4712	TGGGCAA 4718 4761 TGGCCAA 4767	
D	b				
RESULT 8					
US-08-770-315-4					
; Sequence 4, Application US/09770315					
; Patent No. 6429001					
; GENERAL INFORMATION:					
APPLICANT: Chiron Corporation					
TITLE OF INVENTION: Recombinant AAV Packaging Systems					
FILE REFERENCE: 20263-501					
CURRENT APPLICATION NUMBER: US/09/770,315					
PRIORITY FILING DATE: 2001-01-26					
PRIOR APPLICATION NUMBER: US 60/178,536					
PRIOR FILING DATE: 2000-01-26					
NUMBER OF SEQ ID NOS: 8					
SOFTWARE: FastSeq for Windows Version 3.0					
SEQ ID NO 4					
LENGTH: 4072					
TYPE: DNA					
ORGANISM: Unknown					
FEATURE:					
OTHER INFORMATION: recombinant DNA					
US-09-770-315-4					
<hr/>					
Query Match 35.8%; Score 1691.2; DB 4; Length 4072;					
Best Local Similarity 79.1%; Pred. No. 0;					
Matches 2063; Conservative 0; Mismatches 533; Indels 12; Gaps 4;					
Oy			1900	AGCTCCGATGGACTTGTGCCGACAGGTACCAAAACAATGTTCTCGTACGCGGGCATGCT	1959
D	b		1164	AGCTTCATCACTACGACGACAGCGATGCCAAAAACAATGTTCTCGTACCGTGGCATGAA	1223
Oy			1960	TCAGATGCTGTTTCCCTGCAAGACATCGGAGAGAAATGAATCAGAATTTCAACATTTGCTT	2019
D	b		1224	TCTATGCTGTTTCCCTGCAAGACATCGGAGAAATGAATCAGAATTTCAACATTTGCTT	1283
Oy			2020	CAGGCACGGGACGAGACATGTTCAAGAGTGTTCCCGCGGCGGTGACAAATCTCAACCGGT	2079
D	b		1284	CACCTACGGACAGAAAAGACTGTTTAGAGTGCTTCC---CGTGCABAATCTCAACCCGT	1340
Oy			2080	--CGTCAGAAAGAGACGTATCGGAAACTGTGTCATTATCATCTGCTGGGGCGGCG	2136
D	b		1341	TTCTGTGTCGTCAAAAGCGATACGAAACTGTGCTATCATTCATATCATATGGGAAAGT	1400
Oy			2137	TTCCGGAATGCTTGCTCGGCTCGGCGCATCTGTCACAGTGAAGTGAAGTGAATGCTGTTTC	2196
D	b		1401	GCCAGGAA--CGCTTGCACTGCTGCGATCTGCTCAATGTGGATTTGGATGACTGCATCTT	1457

OY	2137	CGAGCAATPAATGAAGCTTAAACAGGATATGCGCTCGGATGTTATCTTCAGATTGGCGTC	2250
Db	1458	TGAACAAATPAATGATTTAATCAAGGTATGGCTCGGATGGTTATCTTCAGATTGGCGTC	1517
OY	2257	AGGACAACTCTCTCTAGGGGCAATTCGCGAGTGGTGGACTTGGAACCTTGGAGCCCGAAGC	2316
Db	1518	AGGACACTCTCTCTAGAGGAATPAACAGTGGTGGAACTCAAACTTGGGCCCAACACAC	1577
OY	2317	CCAAAGCCAAACGCAAAAAGCAGACGACCGCGGGGTTCTGGTGTCTTGGCTACAAAT	2376
Db	1578	CAAAACCCGACAGAGGCGCATAAAGACGACAGCGGGGTTCTTGTGCTTCTGGGTACAAAT	1637
OY	2377	ACCTTGGAGCCCTTCAACGGGACTTCCACAGGGGGAGCCCGCTCAACGCGCGGACGCGAGG	2436
Db	1638	ACCTTGGAGCCCTTCAACGGGACTTCCACAGGGGAGCCCGCTCAACGCGCGGACGCGAGG	1697
OY	2437	CCCTTGCAGCAGCAGAAAGGCTTACGACGACGAGCTCAAAAGCGGGTGCATCATCGTACCTCG	2496
Db	1658	CCCTTGCAGCAGCAGAAAGGCTTACGACGCGGAGCTTGAACGCGGAGACAAACCGTACCTCA	1757
OY	2497	GGTATTAACCAACGCGCGACCGCGAGTTTCAGAGCCGTCTGCAAGAAATACGTCTTTTGGGG	2556
Db	1758	AGTACAAACCAACGCGCGAGGTTTCAAGAGCCGCTTAAAGAAATACGTCTTTTGGGG	1817
OY	2557	GCAACCTTGGGCGAGCAGTCTTCCAGGCGCAAGAGCGGGTCTCGAACCCTTCGGTCTGG	2616
Db	1818	GCAACCTTGGGCGAGCAGTCTTCCAGGCGGAAAMAAGGGTCTTGAACCTTCGGGCGCTGG	1877
OY	2617	TTGAGAGAAAGCGCTTAAGACCGGCTCTGGAAAGAAACGTCCGGTATGAGCAGTCCGCAACAG	2676
Db	1878	TTGAGAGAAACGTCTTAAGACCGGCTCCGGGAAAAAAGAGCCCGGTATGAGCACTCTCTGTGG	1937
OY	2677	AGCCAGACTCTCTCTCGGAGCATTCGGCAAGACAGGCGACAGCCCGCTAAAAAGAGACTCA	2736
Db	1938	AGCCAGACTCTCTCTCGGAGAACCGGAAAGCGGGCGACAGCTGCAAGAAAAAGATTGA	1997
OY	2737	ATTTTGGTCAGACTTGGCGCATCTCAGAGTCAATGCTCCGATCCCAACCTTCGGAAGACCTC	2796
Db	1998	ATTTTGGTCAGACTTGGGACCGCAGACTCAGTACGTAACCCGACGCTCTCGGACAGCGAC	2057
OY	2797	CAGCAACCCCGCGCTGTGGGAACTTACATCAATGGCTTCAAGGGGTGGCGCACCAATGG	2856
Db	2058	CAGCAGCCCCCTCTGTGTCTGGGAACTTAATGATGGCTTACAGGCAATGGCGACCAACATGG	2117
OY	2857	CAGCAATPAAGAAAGGCGCCGACGAGGTGGGTAATGCTCTCAGAGAAATTTGGCATTTGCAAT	2916
Db	2118	CAGCAATPAAGAAAGGCGCCGACGAGGTGGGTAATGCTCTCGGAAATTTGGCATTTGCAAT	2177
OY	2917	CCACATGGCTGGGCGACAGAGTCATCACACCGACCCCGACCTTGGGCTTGGCCACT	2976
Db	2178	CCACATGATGGGCGACAGAGTCATCACACCGACCCCGAACCCTTGGGCGCTTGGCCACT	2237
OY	2977	ACAATPAACCACTCTACAGAGCAATTCACATGCTTCAACGGGGGCCAGCAACGACCAAC	3036
Db	2238	ACAACACCACTCTACAAACAAATTTCCACGCATCA--GAGAGCTTCGAACGACCAATC	2294
OY	3037	ACTACTTCGGCTACAGCAACCCCTGGGGGTATTTTGAATTTCAACAGATTCACATGCACT	3096
Db	2295	ACTACTTTGGCTACAGCAACCCCTGGGGGTATTTTGAATTTCAACAGATTCACATGCACT	2354
OY	3097	TTTTCACACGCTGACTGGGAGCGACTCATCAACCAATTTGGGATTCGCGCCAGAGAC	3156
Db	2355	TTTTCACACGCTGACTGGGAGAGCTCATCAACCAATTTGGGATTCGAGCCCAAGAGAC	2414
OY	3157	TCAACTTCAAACTCTTCAACATCCAAAGTCAAGAGTCAACGAAATGATGGGTCAACA	3216
Db	2415	TCAACTTCAAGCTCTTAAACATTCAAAGTCAAGAGTCAACGAAATGATGGGTCAACGA	2474
OY	3217	CCATCGCTAAATACCTTACAGCAACGCTTCAAGCTTCTCGGACTCGGAGTACAGGCTTC	3276
Db	2475	CGATTTGCCAATPAACCTTACAGCAACGCTTCAAGCTTCTCGGACTCGGAGTACAGGCTTC	2534
OY	3277	CGTACGTCCTCGGCTTCGCGACCAAGGGGCTCTCCCTCCGTTCCCGGCGGACGTGTTC	3336

Db	2535	CGTAGCCTCGGCTCGGGGCAATGAATGCTTCGCGCGTTCCAGCAAGCGTTTCA	2594
Qy	3337	TGATTCGCCAATACGGGCTACCGTCAACGCTCAACATGAGGACCAAGCCGTGGACCTTCAT	3396
Db	2595	TGGTCCACAGATATGATATCTCAACCCCTGMAACAGGGAGTCAAGGCAATGAGACGCTCTT	2655
Qy	3397	CCCTTTTACTGCTCGGAAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACTT	3455
Db	2655	CATTTTACTGCTCGGAGTACTTCTTCTCAGATGCTGATCGGAAACAATTACTT	2714
Qy	3457	TCAGCTAACCTTTTGAAGAGTGCCTTTCCACAGAGCTACGGCGACAGCCAGACCTGG	3516
Db	2715	TCAGCTAACCTTTTGAAGAGCGTTCCCTTTCCACAGAGCTACGCTCAGCCAGACGCTCGG	2774
Qy	3517	ACCGGCTGATGAATCCTCTCATCGACCAATACCTGTATTTACCTGAACGAACTCAAAATC	3576
Db	2775	ACCGTCTATGAATCTCTCATCGACCAAGTACTGTATTTACTTGAAGCAAAACAATC	2833
Qy	3577	AGTCCGGAAGTGCCTCAAAACAAGACTTGTCTGTATTAGCCGTGGGCTTCAGCTGCAATG	3636
Db	2835	CAAGTGAACCAACACGAGTCAAGGCTTCAGTTTCTCAAGCCGGAAGAGTGAATTC	2894
Qy	3637	CTGTTACGCCCAAAAACCTGGCTACTGGAACCTGTTATCGGACGACGGCGTTTAAAA	3696
Db	2895	GGGACCAATCTAGGAACCTGGCTCTCTGACCTGTTCACCGCAGACGAGATCAAAAGA	2955
Qy	3697	CAAAAACAGACAAACAAGCAATTTTACCTGGAACGTGGCTCAAAATATTAACCTCA	3756
Db	2955	CATCTCGGATTAACAACAAGTGAATCTGTGGACTTGAGCTTCAAGTATCACTTCA	3014
Qy	3757	ATGGGCGTGAATCCATCATCAACCTGCGACTGCTATGCGCTCAACAACAAGACGAAG	3814
Db	3015	ATGGCAGAGACTCTGTGGTGAATCCGGGCCCGGCATGGAAGCAACAAGACGATGAAG	3074
Qy	3817	ACAAATCTTTTCCATGAGCGGTGCATGATTTTGGAAAAGAGCGCCGAGCTTCAA	3876
Db	3075	AAAATTTTTTCTCAGACGCGGGTTCATCTTTTGGGAACAAGGCTCAAGAAAACAA	3134
Qy	3877	ACACTGCATTGGAACAATGTCATGATTAAGAAGAAAGAAATTAAGCCCTAACCCCTG	3936
Db	3135	ATGTGCACTTGAAGAAAGTCAATGATTAAGAAGAAAGAAATCAAGCAACAATCCG	3194
Qy	3937	TGGCACCAGAAAGATTGGGACCGTGGCAGTCAATTTCCAGAGCAGCAGACGACCTG	3996
Db	3195	TGGCTAGGAGACAGATATGGTCTGTGATTTACCAACCTCCAGAGGACAACAGACAGAG	3255
Qy	3997	CGACCCGAGATGTGCATCTATGGAGACTTAACCTGCGATGTGTGGCAAGATAGAGCG	4056
Db	3255	CTACCGCAGATGTCAACAACAAGGCGTTCTTCAGAGCATGCTCGGACGAGACAGATG	3314
Qy	4057	TGTAACCTGCAAGGTGCCATTTGGGCCAAAAATCTCTCACACAGATGGAACCTTACCCGT	4116
Db	3315	TGTAACCTCAAGGGGCCCATCTGGGCAAAAGATTCAACAACGAGACGGAATTTTACCCCT	3374
Qy	4117	CTCCTCTTATGGGCGGCTTTGGAATCAAGAACCGGCTCTCAGATCCTCATCAAAAACA	4176
Db	3375	CTCCCTCTATGGGTGGAATTCGGAATTAAACACCTCTCCACAGATTTCTCATGAAGAA	3434
Qy	4177	CGCCTGTTCTGCAATCTCCGGCGGAATTTAGCTCAAAAGTTTGCTTCATTATCA	4236
Db	3435	CCCGGTAATCGCAATCTCTTCGACACACTTCAGTGGCGGCAAGTTTGCTTCCTTACCA	3494
Qy	4237	CCCAATATCTCAACGACCAATGATGTGGAATAATTGAATGGGACCTGCAAGAAAGAAAA	4296
Db	3495	CACAGTATCTCAACGACAGGTGACGGTGGAGATCGATGGAGCTGCAAGAAAGAAAAA	3554
Qy	4297	GCAAGCGCTGAATCCCGAAGTGAAGTACATCAATTTATGCAAAATCTCTCAACGTTG	4356
Db	3555	GCAAGCGTGAATCCCGAATTTCAGTACACTTCCAACTAACAAAGCTGTTAATGTGG	3614
Qy	4357	ATTTTACTGTGGAACAATGAGCTTTTATCTGAGCCTCGCCCATTTGGACCCGTTACC	4416

Db	3615	ACTTTACTGTGGAGACCTAATGAGCGCGGTGATTCAGAGACCTCGCGCCCATTTGGACACACATATAC	3674
Qy	4417	TTACCCCGCCCTCGTAATTTACGTGTATTCAAATAACCGGTGATCGTCTTCACAGTTAC	4476
Db	3675	TGACTCGTAATCTGTAATTCGCTGTGTAATCAATTAACCTTTAATCGCTTCACGTTGAC	3734
Qy	4477	TTTGGTCTCTGTCCTTCTTAATCTATC	4504
Db	3735	TTTGGTCTCTGCGAATTCCTTCTTATC	3762

RESULT 9

US-09-532-594B-3
; Sequence 3, Application US/09532594B

GENERAL INFORMATION:

APPLICANT: Kotin, Robert M.

APPLICANT: Davidson, Beverly

FILE REFERENCE: 14014.0252U2

CURRENT FILING DATE: 2000-03-22

; SOFTWARE: FastSEQ for Windows Ver. 1.0.0.1

LENGTH: 1872

ORGANISM: Artificial Sequence

OTHER INFORMATION

NAME/KEY: CLE

NAME/KEY: misc feature

US-09-532-594B-3

Query Match	30.68;
Post 10001	30.68;

Matches 1620; Conservative 0; Mismatches 249; Indels 6; Gaps

335 ATGCGGGCTTCTACGAGATCGTATCA

Db 1 ATGCCGGGTTTC

395 GGCAITTCGAC
 QY

D6 61 GGCATTTCGACTCTTTTGTGAGCTGGGTGGCCGAC

QY 455 TC 11

Query Match 30.6%; Score 1444.8; DB 4; Length 1872;

OY	335	ATGCGGGGCTTCTAGAGATCGTGAATCAAGGTCGAGGAGCACTTGAGAGAGCATCTGCG	394
Db	1	ATGCGGGGGTTCTAGAGATCGTGTGAAGGTGCCAAGGACCTTGAGACGAGCATCTGCC	60
OY	395	GGCATTTTCTGACTCGTTTGTGAGCTGGTGGTGGCCGAGAAATGGGAAGCTGCCCGGAT	454
Db	61	GGCATTTTCTGACTCTTTTGTGAGCTGGTGGTGGCCGAGAAATGGGAAGCTGCCCGGAT	120
OY	455	TTTGAATGGAATCTGAATCTGAATTGAGCAGGACCCCTGACCGTGGCCGAGAAAGTTGGAG	514
Db	121	TTTGCATAGCAATTGAATCTGAATTGAGCAGGACCCCTGACCGTGGCCGAGAAAGTTGCA	180
OY	515	CGCGACTTCCTGGTGTCAATGGCGCGCGCGTGAATTAAGGCCCGCGAAGCCCTCTTTGTT	574
Db	181	CGCGACTTCCTGGTGTGAGTGGCGCGCGGTGAATTAAGGCCCGCGAAGCCCTCTTTTGTTC	240
OY	575	CAGTTCGAGAAAGGGCGAGTCTACTTCCACTCCATTAATCTGTGTGAGACCAAGCGGAGTTC	634
Db	241	CAGTTCGAGAAAGGGGGAACGCTACTTCCACTCTGCACATCTTGTGTGAGACCGTGGGCGTTC	300
OY	635	AAATCCATGTGTCTGGGCGGCTTCCGTGATCAGATTAAGGACCAAGCTGGTGCAGACATC	694
Db	301	AAATCCATGTGTGTGGGCGGCTGACGTGAGCAGATTAAAGACAACTGGTGAACCCGATTC	360
OY	695	TACCGCGGAGATGAGCCGACCTGTGCCCAACTGTGTTGCGGTGATCCAAAGACGCTAATGGC	754
Db	361	TACCGCGGGGTGAGCCGACGCTTCCGAACGTGTTGCGGTGATCCAAAGACGCTAATGGC	420
OY	755	GCCGGAAGGGGGAACAAGTGTGTGACGAGTCTACATCCCAACTACCTCCTGGCCGAG	814

Db	421	GC	CGAGGACGGGAAACAAGGTGTGTGACGACTGTCAATCCCCAACTACCTGCCCAAG	480
Qy	815	ACT	CAGCCCGCAGCTGCAGTGTGGGGGTGACATTAATGAGAGATTAATAACCGCTTTGG	874
Db	481	ACC	CAGGCCCAAGCTCCAGTGGGGGTGACATTAACAGATTAATAAACCGCTTTGGT	540
Qy	875	AAC	TGGCCGAGCGCAACCGAGCTGTGTGGCGCAGACCTGACCCACGTCAACAGCCAG	934
Db	541	AAT	TCGCGGAGGTAAACGGCTGTGTGGCGCAGATTTGACGACGTGTGGCAGACCGAG	600
Qy	935	GAG	CAGAAACAAGAGAACTTGAACCCCAATTTGACGCGCTGTCAATCCGGTCAAAAAC	994
Db	601	GAG	CAGAAACAAGAGAAACCAAGAACCCCAATTTGACGCGCGGTCAATCAAGTCAAAAAC	660
Qy	995	TCG	CGCGGCTACATGAGAGCTGTGTGGGTGGCTGTGACCGGGGCATCACTCCGAGAG	1054
Db	661	TCG	CCCGAGGTAAATGAGAGCTGTGTGGGTGGCTGTGACCGGGGATCACTCAAGAAAG	720
Qy	1055	CAG	TGATTCAGAGGACCAAGGCGCTGTATCACTCTTCAACGCGCGTTCCAATCCGCG	1114
Db	721	CAAT	TGATTCAGAGGACCAAGGCGCTTCACTCTCTTCAACGCGCGCTTCAACTCCGCG	780
Qy	1115	TC	CCAGATCAAGGCGCGCTGTGACAAATGCCGACAGATCAATGGCGCTGACCAATCCGCG	1174
Db	781	TC	CAATCAATCAAGGCGCGCTGTGACAAATGCCCTCCAAATCAATGACCTGACAAACGGCT	840
Qy	1175	CCG	CATCACTCTGTGTAGGCGCCCGCTCCGCGCGCGGACATTTAAACCAACCGCATTAACCG	1234
Db	841	CCG	CATCACTCTGTGTGGGCGCAAAACCCCGCGAGGACATTTCCAGCAACCGCATTTACCGA	900
Qy	1235	ATC	CTGTGAGCTGAACGGCTTCGAACCTGACCGCGGCTCCGCTTCTTCGCGTGGAGCC	1294
Db	901	ATC	CTGTGAGATGAACGGGTTACGATCCGCAATGCGGCGCTTCGCTTCTTCGCGCTGGGCG	960
Qy	1295	CAG	AAAGGTTCCGGGAAGCGCAACCACTGTGCTGTTTGGGCCGACCAACCGGAGCAAG	1354
Db	961	CAAA	GAAGTTCCGGGAAGAGGAACACATCTGCTTTTGGGCCGACCAAGCGGATAA	1020
Qy	1355	ACC	CAATCCGGGAGCCATTCGCCACGCGGTCTTTCAACGCGTGCCTTCATCGGCTGCCTCAATGAGAC	1414
Db	1021	ACC	CAATCCGGGAGGCGATTCGCCACGCGGTCTTTCAACGCGCTGCCTTCATCGGCTGCCTCAATGAGAC	1080
Qy	1415	AAT	GAGAACTTCCCTTCANATGATTTGGTGTGACAAAGATGGATCTGTGGGAGAGAGGC	1474
Db	1081	AAT	GAGAACTTCCCTTCACGATTTGGTGTGACAAAGATGGATCTGTGGGAGAGAGGC	1140
Qy	1475	AAG	ATGACCGGCAAGCTGTGTAGTCCGCCAAGGCGAATTCCTCGCGGCGAGCAAGTGCGC	1534
Db	1141	AAG	ATGACCGGCAAGCTGTGTAGAGGCGCAAGGCGAATTCCTCGCGGCGAGCAAGTGCGC	1200
Qy	1535	GTG	ACCAAAATGTCANAGTGTGTCCGCCAGATGCAACCCACCCCGTGTATGTCACTCC	1594
Db	1201	GTG	ACCAAAATGTCANAGTGTGTCCGCCAGATGCAACCCACCCCGTGTATGTCACTCC	1260
Qy	1595	AAC	CAAAATGTGGCGCGGTGATTTGAGGGGAACGACCACTTCGAGACACAGCACGCG	1654
Db	1261	AAC	CAAAATGTGGCGCGGTATGAGGGAATTCGACACTTCCTGAGACACAAACCA	1320
Qy	1655	TTG	CAGGACCGGATGTTCAATTTGAACTCAACCGCGCTGTGAGCATGACTTTGGCAAG	1714
Db	1321	CTC	CAGGACCGGATGTTCAAGTTGAGCTCAACAGGCGCTGTGAGCAAGACTTTGGCAAG	1380
Qy	1715	GTG	CAAAAGCAAGATCAAGATTTCTTCGCTGTGGCGCGAGATCAAGTACCAGAGGTG	1774
Db	1381	GT	CAAAAGCAAGATCAAGACTTTTTCGCGGTGGCGGTACAGATCAAGTACCAGAGGTG	1440
Qy	1775	GCG	CATGATTTACGTTCAGAAAAGGTGTGAGCCAAACAAAAGACCCCGCCCGATGAGAGCG	1834
Db	1441	ACT	CAACAGTTTACGTTCAGAAAAGGTGTGAGCTTGAAGAGGCGCCGCCCAATGAGCA	1500
Qy	1835	GAT	TAAAGCGAGCCCAAGCGGCGCTGCCTCACTCGCGATTCATGACGTCAAGCGCG	1894
Db	1501	GAT	TAAAGTGAGCCCAAGCGGCGCTGTTCCTCACTTGTGCGAGCATGAGCTCAAGCGCG	1560

QY 1895 GAAGGAGCTCCGGTGAATTGGCGACAGGTAACAAACAAATGTTCTGCTCAAGCGGCG 1954
Db 1561 GA---AGCTCCGGTGAAGTACGCGACAGGTAACAAACAAATGTTCTGCTCAAGCGGCG 1817
QY 1955 ATGCTTCAGATGCTGTTCCCTCGCAAGACATGCGAGAGAAATGAATCAATTTCAACTT 2014
Db 1618 ATGAATCTGATGCTTTTCCCTCGCGCAATGCGAGAGATGAATCAATGAATGAGCAATT 1677
QY 2015 TCGCTTCACGACGCGAGAGAGACTGTTCAAGAGTCTTCCCCG---GCGTGTCAAAATCT 2071
Db 1678 TCGCTTCACGACGCGAGAGACTGTTCAAGAGTCTTCCCCG---GCGTGTCAAAATCT 1737
QY 2072 CAACCGGTGCTGAGAAAGAGAGATCGGAACTCGTGCATTCATCATCTGCTGCGG 2131
Db 1738 GTCTGTGCTGTCAAGAAAGCGAGATCAAGAAATGTTCTGCTCAATCAATCAATGAGG 1797
QY 2132 CGGCTCTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191
Db 1798 AGGGCGCGCGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
QY 2192 GTTCTGAGCAATTA 2206
Db 1858 GACATGGAACAAATTA 1872

RESULT 10
US-09-532-594B-15
; Sequence 15, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safier, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532.594B
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 Rep 78 gene
; US-09-532-594B-15

Query Match 30.6%; Score 1444.8; DB 4; Length 1872;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 1620; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

Db 241 CAGTTGAGAGAGGAGGAGACAGTACTTCCATCTGCAATCTCTGTTGAGAGACCGTGGCGCTC 300
QY 635 AATTCATGCTGCTGGCGCGCTTCTGAGTCAGATTTAGGACAGAGTGTGACACATTC 694
Db 301 AATTCATGCTGCTGGCGCGCTTCTGAGTCAGATTTAGGAGAGTGTGAGACCGGCTATC 360
QY 695 TACCGGAGAGTGAAGCGGACCGCTGCAATCTGTTGCGGTGACCAAGACGGTAAATGGC 754
Db 361 TACCGGAGAGTGAAGCGGACCGCTGCAATCTGTTGCGGTGACCAAGACGGTAAATGGC 420
QY 755 GCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 814
Db 421 GCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
QY 815 ACTGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 874
Db 481 ACCGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 875 AACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 934
Db 541 AATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 935 GAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 994
Db 601 GAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
QY 995 TCCGCGCGCTACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1054
Db 661 TCCGCGCGCTACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
QY 1055 CAGTGAATCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1114
Db 721 CAATGATCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
QY 1115 TCCAGATCAAGGCGCGCTGAGCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1174
Db 781 TCACAAATCAAGGCGCGCTGAGCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
QY 1175 CCGGACTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1234
Db 841 CCGGACTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
QY 1235 ATCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1294
Db 901 ATCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
QY 1295 CAGAAAGGTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1354
Db 961 CAGAAAGGTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
QY 1355 ACCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1414
Db 1021 ACCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 1415 AATGAGAACTTCCCTTCAATGATTTGCTGAGCAAAATGATTTGCTGAGGAGGAGGAGGAGGAG 1474
Db 1081 AATGAGAACTTCCCTTCAATGATTTGCTGAGCAAAATGATTTGCTGAGGAGGAGGAGGAGGAG 1140
QY 1475 AAGATGACGCGCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1534
Db 1141 AAGATGACGCGCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
QY 1535 GTGAGCAAAAGTGAAGTGTGCTGCGGCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1594
Db 1201 GTGAGCAAAAGTGAAGTGTGCTGCGGCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
QY 1595 AACACCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1654
Db 1261 AACACCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
QY 1655 TTGACAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1714

Db 1321 CTCGAGACCGGATGTTCAAGTTGAGCTCACCAAGCGCTGAGACGACTTTGGCAG 1380
Qy 1715 GTGCAAGACGAGAGTCAAGATCTTCCGCTGGGGGAGATACGATGACCGAGGTG 1774
Db 1381 GTCCACGAGGAGAGTCAAGATCTTCCGCTGGGGGAGATACGATGACCGAGGTG 1440
Qy 1775 GCGCATGAGTCTACGTCAGAAAGGTGAGACCAAAAGACCCGCGCCGATGAGCG 1834
Db 1441 ACTACAGATTTTACGTAGAAAGGTGAGCTGAAAGAGCCGCGCCCAATGAGCGCA 1500
Qy 1835 GATTAAGGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1894
Db 1501 GATTAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Qy 1895 GAAAGAGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1954
Db 1561 GA--AGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1617
Qy 1955 ATGCTTGAATGCTGTTTCCCTGCAAGACATGCAAGATGATGATGATGATGATGAT 2014
Db 1618 ATGATCTGATGCTTTTCCCTGCGGCAATGCAAGATGATGATGATGATGATGAT 1677
Qy 2015 TGCTTCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2071
Db 1678 TGCTTCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1737
Qy 2072 CAACCGGCTGTCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2131
Db 1738 GTGCTGTGTCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1797
Qy 2132 CGGAGCTCCGAGATGCTGCTGCGGCTGCGATCTGTCAGATGTCAGATGTCAGAT 2191
Db 1798 AGGCGCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1857
Qy 2192 GTTCTGAGCATTA 2206
Db 1858 GACATGAGAACATTA 1872

RESULT 11

US-08-331-384-2/c
Sequence 2, Application US/08331384
Patent No. 5856152
GENERAL INFORMATION:
APPLICANT: Wilson, James M.
APPLICANT: Kelley, William M.
APPLICANT: Fisher, Krishna J.
TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and Methods
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/331,384
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: UPNG1149USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4910 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-331-384-2
Query Match 30.2%; Score 1425.8; DB 2; Length 4910;
Best Local Similarity 83.8%; Pred. No. 0;
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;
Qy 205 GTCTGATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 264
Db 2646 GTCTGATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2588
Qy 265 TAGGAT 323
Db 2587 CTGGATATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2528
Qy 324 GAGCAGAGCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 383
Db 2527 GCGAGAGCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2468
Qy 384 AGCAGCTGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 443
Db 2467 GCGATCTGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2408
Qy 444 TGCCCCCGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 503
Db 2407 TGCCCCCGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2348
Qy 504 AGAAGCTGAGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
Db 2347 AGAAGCTGAGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2288
Qy 564 TCTTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623
Db 2287 TTTTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2228
Qy 624 CCAAGGAGGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 683
Db 2227 CCAAGGAGGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2168
Qy 684 TGAGAGACCATCTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 743
Db 2167 TTGAGAGATTTACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2108
Qy 744 CGGATATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 803
Db 2107 CGAAGATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2048
Qy 804 TCTGCGCAAGACTCAGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 863
Db 2047 TGCTCCCGCAAGACCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1988
Qy 864 GCGCTGTTGAACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 923
Db 1987 GCGCTGTTGAACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1928
Qy 924 GCGAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 983
Db 1927 GCGAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1868
Qy 984 GGTCAAAAACCTCGCGGCTCATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1043
Db 1867 GATCAAAAACCTCATGAGT 1808
Qy 1044 CTTCGAGAGAGTGTATCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103
Db 1807 CTTCGAGAGAGTGTATCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1748

1104 CCAACTCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGCGCTGA 1163
1747 CCAACTCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1688
1164 CCAATTCGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1223
1687 CTAACCGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1628
1224 GCATCTACCGCATCCGAGCTGAGCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1283
1627 GGAATTAATAAATTTTGGAACTAAACGGGTACATCCCAATATGCGGCTTCGTTCTTC 1568
1284 TCGGCTGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1343
1567 TGGGATGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1508
1344 CCAAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1403
1507 CTACCGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1448
1404 TCAACTGAGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1463
1447 TAAACTGAGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1388
1464 GAGAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1523
1387 GAGAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1328
1524 GCAAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1583
1327 GCAAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1268
1584 TCGTCACTCCAGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1643
1267 TCGTCACTCCAGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1208
1644 ACCAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1703
1207 ACCAGGAGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1148
1704 ACTTTGGCAAGGTGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1763
1147 ACTTTGGCAAGGTGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1088
1764 TGACCGAGGTGAGCGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1823
1087 TGGTGGAGGTGAGCGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 1028
1824 CCGATGACGCGGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1883
1027 CCGATGACGCGGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 968
1884 CCGTCAAGCGCGGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1943
967 CCGTCAAGCGCGGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 911
1944 GTACCGGAGGTGAGCGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 2003
910 GTACCGGAGGTGAGCGCAATGAGCAATGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 851
2004 ATTTCACATTTGCTTACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2063
850 ATTTCACATTTGCTTACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 794
2064 CAGATCTCAACCGGT---CGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2120
793 CAGATCTCAACCGGT---CGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 734
2121 ATCTGCTGGGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 2180
733 ATCTGCTGGGCGCGCTCCAGATCAAGCGCTCTGAGCAATGCGCGGAGATCAATGAGCTGA 677
2181 TGATGACTGTGTTTGGAGCAATTAATGACTT 2213

Db 676 TGATGACTGTGTTTGGAGCAATTAATGACTT 644

RESULT 12
US-08-836-087-2/c
Sequence 2, Application US/08836087
Patent No. 5871982
GENERAL INFORMATION:
APPLICANT: Trustees of University of Pennsylvania
APPLICANT: Wilson, James M.
APPLICANT: Kelley, William M.
APPLICANT: Fisher, Krishna J.
TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
TITLE OF INVENTION: Methods of Use Thereof
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,087
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/331,384
FILING DATE: 28-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNVN.007PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4910 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-836-087-2

Query Match 30.2%; Score 1425.8; DB 2; Length 4910;
Best local similarity 83.8%; Pred. No. 0;
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

205 GTCCGTAATTAAGTGTGACGAGTGTGCTTTGGACATTTTGGACACGAGGCAATT 264
2646 GTCCGTAATTAAGTGTGACGAGTGTGCTTTGGACATTTTGGACACGAGGCAATT 2588
265 TAGGTAATTAAGTGTGACGAGTGTGCTTTGGACATTTTGGACACGAGGCAATT 323
2587 GTCCGTAATTAAGTGTGACGAGTGTGCTTTGGACATTTTGGACACGAGGCAATT 2528
324 GAGCAGAGCCATGCGGCTTCTACGAGATGTGATCAAGGTGCGGAGGACCTTGAGC 383
2527 GCGCAGCCCGCATGCGGCTTCTACGAGATGTGATCAAGGTGCGGAGGACCTTGAGC 2468
384 AGCAGCTGCGGCGCATTTGACTGCTTTGTGAGCTGGGCTGCGGAGGAAATGGAGC 443
2467 GGCATCTGCGGCGCATTTGACTGCTTTGTGAGCTGGGCTGCGGAGGAAATGGAGC 2408
444 TGCCCCCGAATTTGACATGATCTGATTTGAGAGGACCCCTGACCGTGGCCG 503

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,087
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNPV.007PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4910 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-09-246-320-2

Query Match 30.2%; Score 1425.8; DB 4; Length 4910;
Best Local Similarity 83.8%; Pred. No. 0;
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

QY 205 GTCCTGATTAGCTGTACGCTGAGTCTTTGCGACATTTTGGCACACACGTCGACATT 264
Db 2646 GTCTGTATTAGAGTCACTGAGTGTG-TTTTGCGACATTTTGGCACACACATGTGTGACG 2588
QY 265 TAGGGTATATATGACCGAGTGAAGCAGAGATCTCATTTTG- ACCGCGAAATTGAAAC 323
Db 2587 CTGGGATTTTAAAGCCCAAGTGAAGCAGCAGGATCTCATTTTGAAGGGGAGGTTTGAAAC 2528
QY 324 GACGAGAGCCATGCGCGGCTTCTAGAGATCGATCAAGTCCGAGGAGGACCTGACG 383
Db 2527 GCGGAGCGCCATGCGGAGGTTTACAGATGTGATTAAAGTCCAGAGACCTTGAACG 2468
QY 384 AGCACTGCGCGGACATTTCTGACTCGTTTGTAGCTGGTGGCCGAGAAAGAAATGGAGC 443
Db 2467 GGGATCTGCGCGGACATTTCTGACAGCTTTGTGAATGGTGGCCGAGAAAGAAATGGAGT 2408
QY 444 TGCCTCCGATTTGATGATGATCTGATTTGAGAGGACCCCTGACCGGTGACG 503
Db 2407 TGCCTCCGATTTGATGATGATCTGATTTGAGAGGACCCCTGACCGGTGACG 2348
QY 504 AGAAGCTGACGCGGACATTTCTGCTGCTCAATGCGCGCGCGGATGAAGCCCGGAGGCC 563
Db 2347 AGAAGCTGACGCGGACATTTCTGACGAGATGCGCGCGCGGATGAAGCCCGGAGGCC 2288
QY 564 TCTTCTTTTGTCAAGTGTGAGAGGCGAGTCTACTTCCACTTCCATATTTCTGTGAGA 623
Db 2287 TTTTCTTTTGTCAAGTGTGAGAGGCGAGTCTACTTCCACTTCCATATTTCTGTGAGA 2228
QY 624 CCAAGGGGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 683
Db 2227 CCAAGGGGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2168
QY 684 TGAAGACCATCTACCGCGGATGAGAGCGGACCTCTGCCCACTGCTTCCGCGTGAAGAA 743
Db 2167 TGAAGACCATCTACCGCGGATGAGAGCGGACCTCTGCCCACTGCTTCCGCGTGAAGAA 2108
QY 744 CCGGTAATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 803
Db 2107 CCGGTAATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2048
QY 804 TCTTCCCAAGACTCAAGCCGAGCTGAGTGGCGGTGAGTAAATGAGAGGATATATTA 863
Db 2047 TCTTCCCAAGACTCAAGCCGAGCTGAGTGGCGGTGAGTAAATGAGAGGATATATTA 1988
QY 864 GCGGCTGTTGAACTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 923
Db 1987 GCGGCTGTTGAACTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1928
QY 924 GCCAGACCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 983

Db 1927 CGCAGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1868
QY 984 GGTCAAAAACCTCCGCGCGGTACATGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1043
Db 1867 GATCAAAAACCTCCGCGCGGTACATGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1808
QY 1044 CCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1103
Db 1807 CCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1748
QY 1104 CCAACTCGCGGTCCGAGTCAAGCCGCTTGGACATGCGCGGAGGAGGAGGAGGAGGAGG 1163
Db 1747 CCAACTCGCGGTCCGAGTCAAGCCGCTTGGACATGCGCGGAGGAGGAGGAGGAGGAGG 1688
QY 1164 CCAATTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1223
Db 1687 CTAATAACCGCGCGGTACATGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1628
QY 1224 GCATCTACGCGCATCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1283
Db 1627 GCATCTACGCGCATCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1568
QY 1284 TCGGCTGGGCGCGAGAAAGGTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1343
Db 1567 TGGGATGGGCGAGAAAGGTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1508
QY 1344 CCAAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1403
Db 1507 CTAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1448
QY 1404 TCAACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1463
Db 1447 TAAACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1388
QY 1464 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1523
Db 1387 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1328
QY 1524 GCAAGGTGCGCGTGAACCAAAAGTGAAGTGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 1583
Db 1327 GCAAGGTGCGCGTGAACCAAAAGTGAAGTGTGCGCGGAGGAGGAGGAGGAGGAGGAG 1268
QY 1584 TCGTCACTTCCAAACCAACATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1643
Db 1267 TCGTCACTTCCAAACCAACATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1208
QY 1644 ACCAGAGCGGTTGACGAGACCGGATGTTCAAAATTTGAATCTACCCGCGCTTGAAGCATG 1703
Db 1207 ACCAGAGCGGTTGACGAGACCGGATGTTCAAAATTTGAATCTACCCGCGCTTGAAGCATG 1148
QY 1704 ACTTTGGCAAGGTGACCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1763
Db 1147 ACTTTGGCAAGGTGACCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1088
QY 1764 TGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1823
Db 1087 TGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1028
QY 1824 CCGATACCGCGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1883
Db 1027 CCGATACCGCGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 968
QY 1884 CGTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1943
Db 967 CGTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 911
QY 1944 GTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2003
Db 910 GTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 851
QY 2004 ATTTCAACATTTTGTCTTCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2063

Db 850 ATTGAATATCTGCTTCACTACGAGAGAAAGCTGTTTAGAGCTTTCC---CGTGT 794
Qy 2064 CAGAAATCTCAACCGGT---CGTCAAGAAAGAGAGATGAGAACTGTGTCATTGATC 2120
Db 793 CAGAAATCTCAACCGGTCTGTCTGTCTCAAAAAGAGGATACAGAACTGTGTCATTGATC 734
Qy 2121 ATCTGCTGGGGGGGGGCTCCGAGATGCTTCTGCGGCTGAGATCTGGTCAAGTGGACC 2180
Db 733 ATATCATGAGGAGAAAGGTGACAG---CGTTGCACTGCTGAGATCTGTGATGATTT 677
Qy 2181 TGATGACTGTGTTCTGAGCAATTAATGACTT 2213
Db 676 TGATGACTGTGATCTTTGACATTAATGATTT 644

RESULT 14
US-09-546-738-2/c
Sequence 2, Application US/09546738
Patent No. 6387368
GENERAL INFORMATION:
APPLICANT: Trustees of University of Pennsylvania
Willson, James M.
Kelley, William M.
Fisher, Krishna J.
TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
Methods of Use Thereof
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546,738
FILING DATE: 11-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/246,320
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNVFN.007PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4910 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-546-738-2

Query Match 30.2%; Score 1425.8; DB 4; Length 4910;
Best Local Similarity 83.8%; Pred. No. 0;
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

Qy 205 GTCCGATTAATGCTGACGATGAGTCTTTGGCAGATTGGCAGACCAAGTGGCATT 264
Db 2646 GTCCGATTAATGAGTACGTAAGTGTGCGACATTGGCAGACCAAGTGGTACG 2588
Qy 265 TAGGATATATGCGGAGTGAAGAGAGATCTCCATTGG-ACCGCAATTTGAC 323
Db 265 TAGGATATATGCGGAGTGAAGAGAGATCTCCATTGG-ACCGCAATTTGAC 323

Db 2587 CTGGGATATTAAGCCCGAGTGAAGACGACAGGCTCTCATTTTGAAGCGGAGTTGAAC 2528
Qy 324 GAGACGAGCCATGCGGGCTTTAGAGATCTGATCAAGGTGCGGAGGACTTGACG 363
Db 2527 GCGGAGCGGACGAGCGGGGTTTAAAGATTTGAATTAAGGTCCCGAGGACCTTGACG 2468
Qy 384 AGCACTGCGGGGATTTTGAATCTGATCGTTTGAAGTGGGCGGAGAAAGAAATGGAGC 443
Db 2467 GGCATCTGCGGGGATTTTGAAGCTTTTGAAGTGGGCGGAGAAAGAAATGGAGT 2408
Qy 444 TGCCCCCGAATTTGAATGATCTGAATCTGAATGAATGAAGCAGGACCCCTGACCGTGG 503
Db 2407 TGCGCGAATTTGAATGATCTGAATCTGAATGAATGAATGAAGCAGGACCCCTGACCGTGG 2348
Qy 504 AGAAGCTGCGCGGACTTCTGTGCAATGGGCGCGGTGAATGAAGCCCGGAGGCCC 563
Db 2347 AGAAGCTGCGCGGACTTCTGTGCAATGGGCGCGGTGAATGAAGCCCGGAGGCCC 2288
Qy 564 TCTTCTTTGTTGAGTGAAGAGGCGGCTACTTCCACTTCCATTAATCTGTGTGAGA 623
Db 2287 TTTTCTTTGTTGAGTGAAGAGGCGGCTACTTCCACTTCCATTAATCTGTGTGAGA 2228
Qy 624 CACGCGGAGTCAATTCATGCTGTGGCGCTTCTGAGTCAATTAAGGACAGCTGG 663
Db 2227 CACGCGGAGTCAATTCATGCTGTGGCGCTTCTGAGTCAATTAAGGACAGCTGG 2168
Qy 684 TGCAAGCATCTACCGCGGATGAGCCGACCTGCGCAACTGTGTCGCGTGAACCAAG 743
Db 2167 TTGAAGAAATTTACCGGGATGAGCCGACTTTCGCAAACTGTGCGCGTCAAAAG 2108
Qy 744 CGGTAATGCGCGGAGGAGGAGCAAGGTGTGAGCAAGTCTACATCCCACTACC 803
Db 2107 CCAAGAAATGCGCGGAGGAGGAGCAAGGTGTGAGCAAGTCTACATCCCACTACC 2048
Qy 804 TCTTGCCCAAGTCAAGCCGAGCTGAGTGGCGTGAATTAAGAGATTAATA 863
Db 2047 TGCTCCCAAAACCGACCTGAGCTCAGTGGCGTGAATTAAGAGATTAATA 1988
Qy 864 GCGGCTTTGAACTGCGGAGCGCAAGGCTGCGGCGGAGCACTTGAACCACTGA 923
Db 1987 GCGGCTTTGAACTGCGGAGCGCAAGGCTGCGGCGGAGCACTTGAACCACTGA 1928
Qy 924 GCCAGACCCGAGAGCAAGCAAGATCTGAACCCCAATTTGACGCGCTGTCACTC 983
Db 1927 CGAGAGCGGAGAGCAAGCAAGATCTGAACCCCAATTTGACGCGCTGTCACTC 1868
Qy 984 GGTCAAAACCTCGCGCTGATGAGCTGTGCGGTGCTGTGAGACCGGCGATCA 1043
Db 1867 GATCAAAACCTCGCGCTGATGAGCTGTGCGGTGCTGTGAGACCGGCGATCA 1808
Qy 1044 CTTCCGAGAGCTGATGATCCAGAGAGCCAGGCTGTGATCTCTTCAAGCGGCTT 1103
Db 1807 CTTCCGAGAGCTGATGATCCAGAGAGCCAGGCTGTGATCTCTTCAAGCGGCTT 1748
Qy 1104 CCAACTGCGGCTCCAGATCAAGGCGCTGTGACAATGCGGCAAGATCATGCGCTGA 1163
Db 1747 CCAACTGCGGCTCCAGATCAAGGCGCTGTGACAATGCGGCAAGATCATGCGCTGA 1688
Qy 1164 CCAATTCGCGGCTCCAGATCAAGGCGCTGTGACAATGCGGCAAGATCATGCGCTGA 1223
Db 1687 CCAATTCGCGGCTCCAGATCAAGGCGCTGTGACAATGCGGCAAGATCATGCGCTGA 1628
Qy 1224 GCATCTACCGCATCTGAGAGTGAAGGCTGAGCACTGCTGAGCGGCTGCTTTC 1283
Db 1627 GCATCTACCGCATCTGAGAGTGAAGGCTGAGCACTGCTGAGCGGCTGCTTTC 1568
Qy 1284 TCGGCTGCGGCGGAGAAAGGTGCGGAAAGGCAACATCTGCTGTTTGGCGGCA 1343
Db 1567 TGGGATGCGGCGGAGAAAGGTGCGGAAAGGCAACATCTGCTGTTTGGCGGCA 1508
Qy 1344 CCAAGGCAAGCAATGCGGAGGCAATGCGGCAAGGCGGCTGCTTTCAGGCTGCG 1403
Db 1507 CCAAGGCAAGCAATGCGGAGGCAATGCGGCAAGGCGGCTGCTTTCAGGCTGCG 1448

Db 841 CCGGACTACCTGCTGTTGGGCGAGAACCCGCGGAGGACATTTCCAGCAACCGCATCTACCGA 900
QY 1235 ATCTTGAGCTGAACCGCTACGAACCTGCTACGCCCGGCTCTGTTCTTGCGCTGGGCC 1294
Db 901 ATCTTCGAGATGAACGGGTAGATCCGCAATCGGGCTCCGTTCTTGCGGCTGGGCG 960
QY 1295 CAGAAAAGTTGGGGAAGCGCAACCATCTGGCTGTTGGGCGGCAACCAAGGCGCAAG 1354
Db 961 CAAAGAGATTGGGGAAGAGAAACCATCTGGCTCTTTGGGCGGCGCAAGCGGTAAA 1020
QY 1355 ACCAACAATCGGGAAGCCATCGCCCGCTGCTCTTAAGGCTGCTCACTGAGCC 1414
Db 1021 ACCAACAATCGGGAAGCCATCGCCCGCTGCTCTTAAGGCTGCTCACTGAGCC 1080
QY 1415 AATGAGAACTTCCCTTCAATGATTTGCGTCAAGATGTGATCTGTTGGAGAGAGGC 1474
Db 1081 AATGAGAACTTCCCTTCAAGATTTGCGTCAAGATGTGATCTGTTGGAGAGAGGC 1140
QY 1475 AAGATGACGGCCAGAGTGTGAGATCGCGCAAGGCCATTTCTGGGCGGAGCAAGTGGC 1534
Db 1141 AAGATGACGGCCAGAGTGTGAGAGCGCCAAAGGCCATCTGGGCGGAGCAAGTGGC 1200
QY 1535 GTGAGCCAAAAGTCAAGTCTGCGCCCAAGTCGACCCGCTGATGTCACTCC 1594
Db 1201 GTGAGCCAAAAGTCAAGTCTGCGCCCAAGTCGACCCGCTGATGTCACTCC 1260
QY 1595 AACACCAACATGTGCGCGCTGATTTGACGGGAACAGCACCACTTTCAGACACAGCGG 1654
Db 1261 AACACCAACATGTGCGCGCTGATTTGACGGGAACAGCACCACTTTCAGACACAGCGG 1320
QY 1655 TTGACAGACCGGATGTTCAAAATTTGAACTCACCGCGCTGAGCATGACTTTGGCAAG 1714
Db 1321 CTCCAGACCGGATGTTCAAAATTTGAACTCACCGCGCTGAGCATGACTTTGGCAAG 1380
QY 1715 GTGACAAAAGCAGAAAGTCAAGAGTTCTTCCGCTGGGCGCAGATCACGTGACCGAGGTG 1774
Db 1381 GTGACAAAAGCAGAAAGTCAAGAGTTCTTCCGCTGGGCGCTGAGATCACGTGACCGAGGTG 1440
QY 1775 GCGGATGATCTAGTCAAGAAAGGTGAGCCAAACAAAGACCGCGCCCGGATGAGCGG 1834
Db 1441 ACTCACGAGTTTACGTCAAGAAAGGTGAGCTGAAAGAGCGCCCGCCCAATGACGCA 1500
QY 1835 GATAAAAGCAGACCGCAAGCGGCGCTGCGCTCAATCGCGGATCCATCGACGTCAAGCGG 1894
Db 1501 GATATAAGTGAAGCCCAAGCGGCGCTGCGCTCAATCGCGGATCCATCGACGTCAAGCGG 1560
QY 1895 GAAGAGCTCCGCTGAGACTTTGCCGACAGGT 1925
Db 1561 GA--AGCTCCGCTGAGACTAGCGGACAGAT 1588

Search completed: July 20, 2003, 13:23:48
Job time : 234.626 secs

1.3

Query	Match	Best Local Similarity	100.0%; Score 4718; DB 21; Length 4718;	Matches 4718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	TTGGCCACCTCCCTCTGCGGCGCTGCTGCTCGGCTGGGGGCTGGCGACCAAGGTCGC	60	
DB	1	TTGGCCACCTCCCTCTGCGGCGCTGCTGCTCGGCTGGGGGCTGGCGACCAAGGTCGC	60	
QY	61	AGACGGCAGAGACTGTGCTCTGCGCGGCCCCACCGACGACGACGCGCAGAGAGGAGTG	120	
DB	61	AGACGGCAGAGACTGTGCTCTGCGCGGCCCCACCGACGACGACGCGCAGAGAGGAGTG	120	
QY	121	GGCAATCCCACTATTAGGGGTAATTCGGAACGCGCTCCACGCTGCGCGCGCTGAGGCTGA	180	
DB	121	GGCAATCCCACTATTAGGGGTAATTCGGAACGCGCTCCACGCTGCGCGCGCTGAGGCTGA	180	
QY	181	CGTAATAATTAAGCTCATAGGGGAGTGTCTGTATTAGCTGTACGTAAGTGTCTTTGCGAC	240	
DB	181	CGTAATAATTAAGCTCATAGGGGAGTGTCTGTATTAGCTGTACGTAAGTGTCTTTGCGAC	240	
QY	241	ATTTTGGACACACGCTGGCCATTAGGGTAATATATGCGGAGTGAAGAGCAGATCTC	300	
DB	241	ATTTTGGACACACGCTGGCCATTAGGGTAATATATGCGGAGTGAAGAGCAGATCTC	300	
QY	301	CATTTTGGACCGCGAAATTTGAACGAGCAGCAGCATGCGGGCTCTACAGAGATGAT	360	
DB	301	CATTTTGGACCGCGAAATTTGAACGAGCAGCAGCATGCGGGCTCTACAGAGATGAT	360	
QY	361	CAAGTCCCGACGACCTGACGACACCTCTCGGGCAATTTCTGACTGTTGTAGCTG	420	
DB	361	CAAGTCCCGACGACCTGACGACACCTCTCGGGCAATTTCTGACTGTTGTAGCTG	420	
QY	421	GGTGGCCGAGAGGAATGGAGCTGCGCCCGGATTTGACATGATCTGAATCTGATTGA	480	
DB	421	GGTGGCCGAGAGGAATGGAGCTGCGCCCGGATTTGACATGATCTGAATCTGATTGA	480	

QY 541 CGTAGTAAGAGCCCGGAGGCGCTCTTCTTGTTCAGTTCGAGAAAGGCGAGTCTTACTT 600
DB 541 CGTAGTAAGAGCCCGGAGGCGCTCTTCTTGTTCAGTTCGAGAAAGGCGAGTCTTACTT 600
QY 601 CCACCTTCATATTCGTGTGAGAGCAACGAGGAGTCAATTCATGTGTGTGAGCGCTTCT 660
DB 601 CCACCTTCATATTCGTGTGAGAGCAACGAGGAGTCAATTCATGTGTGTGAGCGCTTCT 660
QY 661 GAGTCAGATTAGAGAGCAAGCTGTGTGAGAGCAATTCATGTGTGTGAGCGAGTCTGAG 720
DB 661 GAGTCAGATTAGAGAGCAAGCTGTGTGAGAGCAATTCATGTGTGTGAGCGAGTCTGAG 720
QY 721 CAATGCTTCCGAGTTCAGAGAGCGCTTTCAGAGAGCGAGAGGAGAGCAAGGTGTGAG 780
DB 721 CAATGCTTCCGAGTTCAGAGAGCGCTTTCAGAGAGCGAGAGGAGAGCAAGGTGTGAG 780
QY 781 CGAGTGTACATCCCACTACCTCTGCGCAGAGCTCAAGCCGAGTGTGAGTGTGAG 840
DB 781 CGAGTGTACATCCCACTACCTCTGCGCAGAGCTCAAGCCGAGTGTGAGTGTGAG 840
QY 841 GACTTAATAGAGAGATATTAAGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
DB 841 GACTTAATAGAGAGATATTAAGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
QY 901 GGGCGAGACCTGACCCAGCTGAGCGAGAGCCAGAGAGAGAGAGAGAGTGTGAG 960
DB 901 GGGCGAGACCTGACCCAGCTGAGCGAGAGCCAGAGAGAGAGAGAGAGTGTGAG 960
QY 961 CAATTCGAGCGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1020
DB 961 CAATTCGAGCGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1020
QY 1021 GTGGCTGTGTGAGCGGGGAGTCACTCTCGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1080
DB 1021 GTGGCTGTGTGAGCGGGGAGTCACTCTCGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1080
QY 1081 GTACATCTCTTCAACGCGCTTTCAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1140
DB 1081 GTACATCTCTTCAACGCGCTTTCAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1140
QY 1141 TGCGGAGAGATCAATGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1200
DB 1141 TGCGGAGAGATCAATGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1200
QY 1201 GCGCGGAGATCAATGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1260
DB 1201 GCGCGGAGATCAATGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1260
QY 1261 TGCGTGTGAGCGGCGTGTGTCTTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1320
DB 1261 TGCGTGTGAGCGGCGTGTGTCTTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1320
QY 1321 CATCTGT 1380
DB 1321 CATCTGT 1380
QY 1381 GCGCGT 1440
DB 1381 GCGCGT 1440
QY 1441 CGTGTGAGAGATGT 1500
DB 1441 CGTGTGAGAGATGT 1500
QY 1501 CGCGAGAGAGATGT 1560
DB 1501 CGCGAGAGAGATGT 1560
QY 1561 CGAGATGAG 1620
DB 1561 CGAGATGAG 1620

QY 1621 CGGAG 1680
DB 1621 CGGAG 1680
QY 1681 ACTGAG 1740
DB 1681 ACTGAG 1740
QY 1741 CTTGCGT 1800
DB 1741 CTTGCGT 1800
QY 1801 TGAG 1860
DB 1801 TGAG 1860
QY 1861 CCGCTGAGT 1920
DB 1861 CCGCTGAGT 1920
QY 1921 CAGTGTGAG 1980
DB 1921 CAGTGTGAG 1980
QY 1981 GACATGAG 2040
DB 1981 GACATGAG 2040
QY 2041 TTGAGAGT 2100
DB 2041 TTGAGAGT 2100
QY 2101 GAAAGT 2160
DB 2101 GAAAGT 2160
QY 2161 CGATCTGT 2220
DB 2161 CGATCTGT 2220
QY 2221 GATGAGT 2280
DB 2221 GATGAGT 2280
QY 2281 GCGAGT 2340
DB 2281 GCGAGT 2340
QY 2341 ACAG 2400
DB 2341 ACAG 2400
QY 2401 ACAAG 2460
DB 2401 ACAAG 2460
QY 2461 ACCAG 2520
DB 2461 ACCAG 2520
QY 2521 TTGAG 2580
DB 2521 TTGAG 2580
QY 2581 AGGAG 2640
DB 2581 AGGAG 2640
QY 2641 CTGAG 2700
DB 2641 CTGAG 2700
QY 2701 GCAAG 2760

D	2701	GCAGACGAGCGAGAGCCCGCTAAAGAGCTCAATTTTGTCTAGCTGGCAGCTAG	2760	D	3781	CTGGCACTGCTATGAGCCTCACACAAGACGACGAGCAAGTCTTTCCATGAGCGGTG	3840
Q	2761	AGTCAATCCCGATCCACACCTCTCGGAGAACTCCAGCAACCCCGCTGCTGGAGC	2820	Q	3841	TCATGATTTTGGAAAAAGAGGCGGAGCTTCAAAACATCGATTTGACAAATGTATGA	3900
D	2761	AGTCAATCCCGATCCACACCTCTCGGAGAACTCCAGCAACCCCGCTGCTGGAGC	2820	D	3841	TCATGATTTTGGAAAAAGAGGCGGAGCTTCAAAACATCGATTTGACAAATGTATGA	3900
Q	2821	CTACACAAATGAGCTTCAAGCGGTGGCGACCAATGGCAGCAATTAAGSAGCGCCGACG	2880	Q	3901	TTACAGACGAGAGAGAAATTAAGACCACTAACCTGTGGCCACCGAAAGATTGGAGCG	3960
D	2821	CTACACAAATGAGCTTCAAGCGGTGGCGACCAATGGCAGCAATTAAGSAGCGCCGACG	2880	D	3901	TTACAGACGAGAGAGAAATTAAGACCACTAACCTGTGGCCACCGAAAGATTGGAGCG	3960
Q	2881	GAGTGGGTATGCTCAGAGAAATTTGGCAATTCGATTCACATGCTGGGCGACAGATGA	2940	Q	3961	TGGCAGTCAATTTCCAGAGCAGCAGCAGACCTTGCGACCGGAGATGTGATGTATGG	4020
D	2881	GAGTGGGTATGCTCAGAGAAATTTGGCAATTCGATTCACATGCTGGGCGACAGATGA	2940	D	3961	TGGCAGTCAATTTCCAGAGCAGCAGCAGACCTTGCGACCGGAGATGTGATGTATGG	4020
Q	2941	TCACCAACGACAGCCCGACCTGGGCTTGGCCCACTTAACAAATACCACTCTACAGCAAA	3000	Q	4021	GAGCATTAACCTGGCATGCTGTGGCAAGTATGAAACGTATCCGAGAGGTCCCATTTGGG	4080
D	2941	TCACCAACGACAGCCCGACCTGGGCTTGGCCCACTTAACAAATACCACTCTACAGCAAA	3000	D	4021	GAGCATTAACCTGGCATGCTGTGGCAAGTATGAAACGTATCCGAGAGGTCCCATTTGGG	4080
Q	3001	TCTCAGTGTTCACACGGGGGCGACAGCAACCACTACTTGGGCTACAGACCCCT	3060	Q	4081	CCAAATTCCTCACACAGATGACACCTTTCACCCGTCCTCTTATGAGCGGCTTTGAC	4140
D	3001	TCTCAGTGTTCACACGGGGGCGACAGCAACCACTACTTGGGCTACAGACCCCT	3060	D	4081	CCAAATTCCTCACACAGATGACACCTTTCACCCGTCCTCTTATGAGCGGCTTTGAC	4140
Q	3061	GGGGGTATTTTGAATTTCAACAGATTCACCTGCACTTTTCAACAGTGACTGGACGAC	3120	Q	4141	TCAGAAACCCGCTCTCTAGATTCCTCAACAAAACAGCGCTGTTCTGCGAATCTCCG	4200
D	3061	GGGGGTATTTTGAATTTCAACAGATTCACCTGCACTTTTCAACAGTGACTGGACGAC	3120	D	4141	TCAGAAACCCGCTCTCTAGATTCCTCAACAAAACAGCGCTGTTCTGCGAATCTCCG	4200
Q	3121	TCATCAACAAATTTGGGAGATTTCCGGCCCAAGACACTTAACCTTTCAACATCC	3180	Q	4201	CGAGTTTACGATCAAAAGTTTGGCTTCATTCACACCCAAATCTCCACAGCAAGTGA	4260
D	3121	TCATCAACAAATTTGGGAGATTTCCGGCCCAAGACACTTAACCTTTCAACATCC	3180	D	4201	CGAGTTTACGATCAAAAGTTTGGCTTCATTCACACCCAAATCTCCACAGCAAGTGA	4260
Q	3181	AAGTCAAGAGAGTCAACGAGATGATGAGTCAACCAACATCGTAAATTAACCTTACAGCA	3240	Q	4261	GTTGTGAAATTTGATGGAGCTGACAGAAAGAAAACAGCAAGCGCTGGAATCCGGAAGTGC	4320
D	3181	AAGTCAAGAGAGTCAACGAGATGATGAGTCAACCAACATCGTAAATTAACCTTACAGCA	3240	D	4261	GTTGTGAAATTTGATGGAGCTGACAGAAAGAAAACAGCAAGCGCTGGAATCCGGAAGTGC	4320
Q	3241	CGGTTCAAGTCTTCTCGGACTCGAGTACCAAGCTTCGTAAGTCTGCGGCTCTGCGAC	3300	Q	4321	AGTACATCATCAATTAATGCAAAATCTGCCAACCTTAATTTACTGTGACAAATGAGAC	4380
D	3241	CGGTTCAAGTCTTCTCGGACTCGAGTACCAAGCTTCGTAAGTCTGCGGCTCTGCGAC	3300	D	4321	AGTACATCATCAATTAATGCAAAATCTGCCAACCTTAATTTACTGTGACAAATGAGAC	4380
Q	3301	AGGGCTGCTCCCTCCGTTCCCGGGGAGCGTTCATGATTCGCCAATACCGCTACCTGGA	3360	Q	4381	TTTATACGAGCTGCGCCCATGAGCAACCGCTTACCTTACCCGCTCCCTGTAATTCGTG	4440
D	3301	AGGGCTGCTCCCTCCGTTCCCGGGGAGCGTTCATGATTCGCCAATACCGCTACCTGGA	3360	D	4381	TTTATACGAGCTGCGCCCATGAGCAACCGCTTACCTTACCCGCTCCCTGTAATTCGTG	4440
Q	3361	CGCTCAACATGAGGACCGGAGCGGTGACCTTATCTTACTGCTGAGAAATTTTCC	3420	Q	4441	TTATCAATTAACCGGTGATGATGTTTCAAGTTGAACTTTGATCCTGCTCTTATCT	4500
D	3361	CGCTCAACATGAGGACCGGAGCGGTGACCTTATCTTACTGCTGAGAAATTTTCC	3420	D	4441	TTATCAATTAACCGGTGATGATGTTTCAAGTTGAACTTTGATCCTGCTCTTATCT	4500
Q	3421	CTTCTCAGATGCTGAGAACGGGCAACCTTTACCTTACGCTACACCTTTGAGGAGTGC	3480	Q	4501	TATCGTTACCATGCTTATGCTTAAACATTAACCTGCTTGGTGGCTTGCATTAAG	4560
D	3421	CTTCTCAGATGCTGAGAACGGGCAACCTTTACCTTACGCTACACCTTTGAGGAGTGC	3480	D	4501	TATCGTTACCATGCTTATGCTTAAACATTAACCTGCTTGGTGGCTTGCATTAAG	4560
Q	3481	CTTTCACAGAGAGTACGCGGACCGGACCGGCTGAGACCTGAGTATGATCTTCTCATTCG	3540	Q	4561	ACTTACGTCATGCGGTTAACCCTAGTATGATGATGCTCCACTCTCTCTGCGCGCTGC	4620
D	3481	CTTTCACAGAGAGTACGCGGACCGGACCGGCTGAGACCTGAGTATGATCTTCTCATTCG	3540	D	4561	ACTTACGTCATGCGGTTAACCCTAGTATGATGATGCTCCACTCTCTCTGCGCGCTGC	4620
Q	3541	ACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAAGTCCGAGAGTCCCAAAACAAGG	3600	Q	4621	TCGCTGCTGGGCGCTGCGGACCAAAAGTCCGCAACGGAAGAGCTCTGCTGCGCGGC	4680
D	3541	ACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAAGTCCGAGAGTCCCAAAACAAGG	3600	D	4621	TCGCTGCTGGGCGCTGCGGACCAAAAGTCCGCAACGGAAGAGCTCTGCTGCGCGGC	4680
Q	3601	ACTTGTGTTTAAAGCGTGGGTCTTCAAGTGGCATGCTTTCAGCCCAAAAATGCGCTAC	3660	Q	4681	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
D	3601	ACTTGTGTTTAAAGCGTGGGTCTTCAAGTGGCATGCTTTCAGCCCAAAAATGCGCTAC	3660	D	4681	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
Q	3661	CTGACACCTGTATTCGAGAGAGCGGCTTTTAAACAAAACAGACAAACAGCA	3720	Q	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
D	3661	CTGACACCTGTATTCGAGAGAGCGGCTTTTAAACAAAACAGACAAACAGCA	3720	Q	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
Q	3721	ATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGGGCTGAATTCATCAACC	3780	Q	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
D	3721	ATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGGGCTGAATTCATCAACC	3780	D	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
Q	3781	CTGGCACTGCTATGAGCCTCACACAAGACGACGAGCAAGTCTTTCCATGAGCGGTG	3840	Q	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718
D	3781	CTGGCACTGCTATGAGCCTCACACAAGACGACGAGCAAGTCTTTCCATGAGCGGTG	3840	D	4718	CCACCGAGCGAGCGGCGGACAGAGGAGTGGGCA	4718

RESULT 2

AAAF23749 standard; DNA; 4683 BP.

AAAF23749;

28-MAR-2001 (first entry)

AAV6 DNA sequence.

KM AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 KM atherosclerosis; sickle cell anemia; thalassemia;
 KM blood clotting disorder; diabetes; ss.
 OS Adeno associated virus.
 XX US6156303-A.
 XX PD 05-DEC-2000.
 XX PF 11-JUN-1997; 97US-0873168.
 XX PR 11-JUN-1997; 97US-0873168.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Russell DW, Rutledge EA;
 XX MPI; 2001-060164/07.
 XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalassemia and diabetes
 XX PS Claim 1; Fig 1; 50pp; English.
 CC The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is the DNA sequence of one such serotype (AAV6). AAV6
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.
 CC
 SQ Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
 Query Match 90.3%; Score 4258; DB 22; Length 4683;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 4474; Conservative 0; Mismatches 205; Indels 43; Gaps 5;

QY 478 TGAGCAGGACCCCTGACCGTGGCCGAGAGCTGACGCGCACTTCTGTCATATGGC 537
 DB 463 TGAGCAGGACCCCTGACCGTGGCCGAGAGCTGACGCGCACTTCTGTCATATGGC 522
 QY 538 CCGCGTGAATAGGCCCCCGAGAGCCCTCTTCTTGTTCAGTTGAGAGGCGAGTCTTA 597
 DB 523 CCGCGTGAATAGGCCCCCGAGAGCCCTCTTCTTGTTCAGTTGAGAGGCGAGTCTTA 582
 QY 598 CTTCCACCTTCATATTTCTGTGAGACCAACGGGGGTCAAAATCATGTGTCTGGCCGCTT 657
 DB 583 CTTCCACCTTCATATTTCTGTGAGACCAACGGGGGTCAAAATCATGTGTCTGGCCGCTT 642
 QY 658 CCGAGTCAATTAAGGAGCAAGCTGGTGAAGACCAATCTACCGCGGATGAGCCGACCT 717
 DB 643 CCGAGTCAATTAAGGAGCAAGCTGGTGAAGACCAATCTACCGCGGATGAGCCGACCT 702
 QY 718 GCCCAACTGATTCGCGGTGACCAAGACGGTATAGCGCGCGGAGGAGGAGCAAGTGTGT 777
 DB 703 GCCCAACTGATTCGCGGTGACCAAGACGGTATAGCGCGCGGAGGAGGAGCAAGTGTGT 762
 QY 778 GAGAGTGTATCATTCCTCACTACTCTCTGCCCCAAGCTCAAGCCGAGCTGAGTGGC 837
 DB 763 GAGAGTGTATCATTCCTCACTACTCTCTGCCCCAAGCTCAAGCCGAGCTGAGTGGC 822
 QY 838 GTGACTAATAGAGAGATATATATAGCGCGCTTGAACCTCGCGGAGCGCAAGCGCT 897
 DB 823 GTGACTAATAGAGAGATATATATAGCGCGCTTGAACCTCGCGGAGCGCAAGCGCT 882
 QY 898 CGTGGCAGACACTGACCACTGACGAGACCCAGAGACCAAGAGAGATCTGA 957
 DB 883 CGTGGCAGACACTGACCACTGACGAGACCCAGAGACCAAGAGAGATCTGA 942
 QY 958 CCCCAATTGAGCGCGCTGTATCCGCTCAAAAACCTTCGCGGCTATAGAGCTGTGT 1017
 DB 943 CCCCAATTGAGCGCGCTGTATCCGCTCAAAAACCTTCGCGGCTATAGAGCTGTGT 1002
 QY 1018 CCGGTGGCTGTGAGACCGGGGACATCACTCCGAGAGAGTGGATCCAGAGACCAAGC 1077
 DB 1003 CCGGTGGCTGTGAGACCGGGGACATCACTCCGAGAGAGTGGATCCAGAGACCAAGC 1062
 QY 1078 CTGTACATCTCTTCAAGCGCGCTTCAACTGCGAGTCCAGATCAAGGCGCTGTGA 1137
 DB 1063 CTGTACATCTCTTCAAGCGCGCTTCAACTGCGAGTCCAGATCAAGGCGCTGTGA 1122
 QY 1138 CAATGCGGCAAGATCATGCGCTGACCAATTCGCGCTCACTACCTGTAGGCGCTGC 1197
 DB 1123 CAATGCGGCAAGATCATGCGCTGACCAATTCGCGCTCACTACCTGTAGGCGCTGC 1182
 QY 1198 TCGGCGCGGAGACATTAAACCAACCGCATCTACCGCATCTGTAGAGTGAACGAGCTGCA 1257
 DB 1183 TCGGCGCGGAGACATTAAACCAACCGCATCTGTAGAGTGAACGAGCTGCA 1242
 QY 1258 ACCTGCTAGCGCGGCTCCGCTCTTCTCGGCTGAGGCGCAAAAAGGTTGAGAAAGCGCA 1317
 DB 1243 ACCTGCTAGCGCGGCTCCGCTCTTCTCGGCTGAGGCGCAAAAAGGTTGAGAAAGCGCA 1302
 QY 1318 CACCATCTGAGCTGTGTGAGCGCGCAACCGGCAAGACCAATCGCGGAAAGCATCGC 1377
 DB 1303 CACCATCTGAGCTGTGTGAGCGCGCAACCGGCAAGACCAATCGCGGAAAGCATCGC 1362
 QY 1378 CCAAGCGGTGCGCTTCAAGCGGTGAGTCACTGAGACCAATGAGAACTTTCCCTTCAATGA 1437
 DB 1363 CCAAGCGGTGCGCTTCAAGCGGTGAGTCACTGAGACCAATGAGAACTTTCCCTTCAATGA 1422
 QY 1438 TTGCGTGAACAAGATGTATCTGTGAGAGAGGCAAGATGACGCGCAAGTCTGTGA 1497
 DB 1423 TTGCGTGAACAAGATGTATCTGTGAGAGAGGCAAGATGACGCGCAAGTCTGTGA 1482
 QY 1498 GTCCGCAAGGCGCATTTCTGCGCGGAGCAAGGTGCGGTGACCAAAAGTGAAGTCTGT 1557
 DB 1483 GTCCGCAAGGCGCATTTCTGCGCGGAGCAAGGTGCGGTGACCAAAAGTGAAGTCTGT 1542

1558 CGCCAGATCGACCCCAACCCCGGTGATGTCACCTCCAAACCAACATGTCGCGGTAT 1617
1543 CGCCAGATCGATCCACCCCGGTGATGTCACCTCCAAACCAACATGTCGCGGTAT 1602
1618 TGACGGGAACGACCACTTCGAGACACGACGCGGTGCAAGACCGGATGTTCAATT 1677
1603 TGAAGGGAACGACCACTTCGAGACACGACGCGGTGCAAGACCGGATGTTCAATT 1662
1678 TGAATCAACCGCGGTGAGACATGACTTTGGAGAGTGACAAAGGAGGAAAGTCAAGA 1737
1663 TGAATCAACCGCGGTGAGACATGACTTTGGAGAGTGACAAAGGAGGAAAGTCAAGA 1722
1738 GTTCTTCGCTGGGCGAGATCACTGACGAGGTGGCGCATGATGTTCTACGTCAGAAA 1797
1723 GTTCTTCGCTGGGCGAGATCACTGACGAGGTGGCGCATGATGTTCTACGTCAGAAA 1782
1798 GGGTGGAGCCAAAGAGACCCCGCGGTGAGAGCGGATTAAGGCGGACCGGCG 1857
1783 GGGTGGAGCCAAAGAGACCCCGCGGTGAGAGCGGATTAAGGCGGACCGGCG 1842
1858 CTGCGCCCTCAGTCCGCGATTCATGACGTCAGAGCGGAGAGAGTCCGCTGCACTTGC 1917
1843 CTGCGCCCTCAGTCCGCGATTCATGACGTCAGAGCGGAGAGTCCGCTGCACTTGC 1902
1918 CGACAGGTACCAAAACAAATGTTCTGTCAGCGGCGCATGCTGAGATGCTGTTCCCTG 1977
1903 CGACAGGTACCAAAACAAATGTTCTGTCAGCGGCGCATGCTGAGATGCTGTTCCCTG 1962
1978 CAAGCATGCGAGAGATGAATCAGAAATTTCAACTTTGCTTCAAGCAAGGAGAGAGA 2037
1963 CAAGCATGCGAGAGATGAATCAGAAATTTCAACTTTGCTTCAAGCAAGGAGAGAGA 2022
2038 CTGTTCAAGAGTCTCCCGGCGGTGTCAGAAATCTCAACCGGTCGTCAAGAAAGAGTCTA 2097
2023 CTGTTCAAGAGTCTCCCGGCGGTGTCAGAAATCTCAACCGGTCGTCAAGAAAGAGTCTA 2082
2098 TCGGAACCTGTGCGCATTCATCATCTGCTGGGCGGCGCTCCGAGATTGCTTGGCGC 2157
2083 TCGGAACCTGTGCGCATTCATCATCTGCTGGGCGGCGCTCCGAGATTGCTTGGCGC 2142
2158 CTGCGATGTGTCACGTCGAGACCTGAGATGACTGTGTTCTGAGCAATTAAGACTTAAC 2217
2143 CTGCGATGTGTCACGTCGAGATGACTGTGTTCTGAGCAATTAAGACTTAAC 2202
2218 CAGGATGCTGCGCATGCTTATCTTCCAGATGTCGAGAGCAACCTCTCTGAGGCA 2277
2203 CAGGATGCTGCGCATGCTTATCTTCCAGATGTCGAGAGCAACCTCTCTGAGGCA 2262
2278 TTCGCGATGTGAGGACTTGAACCTGAGAGCCCGAAGCCCAAGCCAGCAAGCAAGC 2337
2263 TTCGCGATGTGAGGACTTGAACCTGAGAGCCCGAAGCCCAAGCCAGCAAGCAAGC 2322
2338 AGGACGAGCGCGGCTGCTGCTGCTTCTGCTGACAAATGCTGAGACCTTCAACGAGC 2397
2323 AGGACGAGCGCGGCTGCTGCTTCTGCTGACAAATGCTGAGACCTTCAACGAGC 2382
2398 TCGACAAAGGGGAGCCGTCACAGCGGCGAGACGAGCGGCTCTGAGACGACAGGCT 2457
2383 TCGACAAAGGGGAGCCGTCACAGCGGCGAGACGAGCGGCTCTGAGACGACAGGCT 2442
2458 ACAGACAGCAGCTCAAGCGGAGTGAACAATCCGTAACCTGCGGTATTAACCAAGCGGCG 2517
2443 ACAGACAGCAGCTCAAGCGGAGTGAACAATCCGTAACCTGCGGTATTAACCAAGCGGCG 2502
2518 AGTTTCAGGAGCTGTCAGAAAGTATGCTTTTGGGGGCAACTCGGGGAGAGTCT 2577
2503 AGTTTCAGGAGCTGTCAGAAAGTATGCTTTTGGGGGCAACTCGGGGAGAGTCT 2562
2578 TCGACGCGCAAGAGGGGTTCTCGAACTTTTGGTCTGTTGAAGAGAGTCTAAGACG 2637
2563 TCGACGCGCAAGAGGGGTTCTCGAACTTTTGGTCTGTTGAAGAGAGTCTAAGACG 2622
2638 CTCTGGAAGAAAGTCGCGGTGAGAGTGCACCAAGAGCAGACTCTCTCGGCGCA 2697

2623 CTCTGGAAGAAAGTCGCGGTGAGAGTGCACCAAGAGCAGACTCTCTCGGCGCA 2682
2698 TCGGCAAGAGAGCGCGAGCGGCTTAAAGAGACTCAATTTTGGTCACTGGGAGCT 2757
2683 TTGGCAAGAGAGCGCGAGCGGCTTAAAGAGACTCAATTTTGGTCACTGGGAGCT 2742
2758 CAGAGTCAGTCCCGGATCCCAACTCTCTGAGAACTCCAGAACCCCGCTGCTGTGG 2817
2743 CAGAGTCAGTCCCGGATCCCAACTCTCTGAGAACTCCAGAACCCCGCTGCTGTGG 2802
2818 GACCTACTAATGCTTCAAGCGGTGGCGACCAATGCGACCAATTAAGAAAGCGCG 2877
2803 GACCTACTAATGCTTCAAGCGGTGGCGACCAATGCGACCAATTAAGAAAGCGCG 2862
2878 ACGAGTGGGTAAATGCTCAAGAAATTTGGCATTTGCAATGCGATGGGCGAGAG 2937
2863 ACGAGTGGGTAAATGCTCAAGAAATTTGGCATTTGCAATGCGATGGGCGAGAG 2922
2938 TCATCAACACAGACACCGGAGCTGGGCTTGGCCACTTAACAAATCAACTCTTACAGC 2997
2923 TCATCAACACAGACACCGGAGCATGGGCTTGGCCACTTAACAAATCAACTCTTACAGC 2982
2998 AAATCTCAGTCTTCAAGCGGCGGCGAGACAGCAACCACTTCTGCTTACAGACCC 3057
2983 AAATCTCAGTCTTCAAGCGGCGGCGAGACAGCAACCACTTCTGCTTACAGACCC 3042
3058 CTTGGGGGTATTTGATTTCAAGATTCACAGTCCGCACTTTTACACAGGATGGGAGC 3117
3043 CTTGGGGGTATTTGATTTCAAGATTCACAGTCCGCACTTTTACACAGGATGGGAGC 3102
3118 GACTCATCAACCAACTTGGGAGATTCGCGCCCAAGAGACTCAACTTCAACTCTTCAACA 3177
3103 GACTCATCAACCAACTTGGGAGATTCGCGCCCAAGAGACTCAACTTCAACTCTTCAACA 3162
3178 TCGAATCAAGAGGTCACAGCAATGATGGGCTCACACCAATGCTTAATTAACCTTACCA 3237
3163 TCGAATCAAGAGGTCACAGCAATGATGGGCTCACACCAATGCTTAATTAACCTTACCA 3222
3238 GCAAGTTCAAGTCTCTGAGACTCGGAGTACAGACTTCCGTAAGTCTGCGCTGCGC 3297
3223 GCAAGTTCAAGTCTCTGAGACTCGGAGTACAGACTTCCGTAAGTCTGCGCTGCGC 3282
3298 ACCAGGCTGCTCTCCCTGCTCCGCGGAGCGTTCATGATTCGCAATACGCTTACC 3357
3283 ACCAGGCTGCTCTCCCTGCTCCGCGGAGCGTTCATGATTCGCAATACGCTTACC 3342
3358 TGAAGCTCAACAAATGGAGCAGAGCGGTGGAGTCTTCTTACTGCTGCTGAATAT 3417
3343 TGAAGCTCAACAAATGGAGCAGAGCGGTGGAGTCTTCTTACTGCTGCTGAATAT 3402
3418 TCCCTTCTGAGTGTGAGAAAGGGCAACAACTTAACTTACCTTACCTTGAAGAG 3477
3403 TCCCTTCTGAGTGTGAGAAAGGGCAACAACTTAACTTACCTTACCTTGAAGAG 3462
3478 TGCCTTTCACAGAGCTAGCGCAAGCGAGAGCTTGAACGCGGTGATTAATCTCTCA 3537
3463 TGCCTTTCACAGAGCTAGCGCAAGCGAGAGCTTGAACGCGGTGATTAATCTCTCA 3522
3538 TCGACCAATACCTGATTAACCTGAACAGAACTCAAAATCAATGCTCCGGAAGTCCCAAAACA 3597
3523 TCGACCAATACCTGATTAACCTGAACAGAACTCAAAATCAATGCTCCGGAAGTCCCAAAACA 3582
3598 AGGACTGCTGTTAGCGGTGCTCAGACTGAGCATGCTGTTTACGCGCCCAAAACTGGC 3657
3583 AGGACTGCTGTTAGCGGTGCTCAGACTGAGCATGCTGTTTACGCGCCCAAAACTGGC 3642
3658 TACCTGAGACCTGTTATCGGACAGCGCGTTTCTAAACAAAAACAGACCAACACACA 3717
3643 TACCTGAGACCTGTTATCGGACAGCGCGTTTCTAAACAAAAACAGACCAACACACA 3702
3718 GCAATTTTACCTGAGCTGCTGCTTCAAAATTAATCACTTGGGCGGATTCATCATCA 3777

```

Db 3703 GCACTTTACCTGAGCTGCTTCAAAATATTAACCTTAATGAGCGTGAATCTATATCA 3762
Qy 3778 ACCCTGGACCTGCTATGCTCTCAACAAGAGCGAAGCAAGTCTTTTCCATGACG 3837
Db 3763 ACCCTGGACCTGCTATGCTCTCAACAAGAGCGAAGCAAGTCTTTTCCATGAGCG 3822
Qy 3838 GTGTGATGATTTTGGAAAAGAGAGCGCGGAGCTTCAAACTGCACTTGGACATATCA 3897
Db 3823 GTGTGATGATTTTGGAAAAGAGAGCGCGGAGCTTCAAACTGCACTTGGACATATCA 3882
Qy 3898 TGATTCAGACGAGAGAGAAATTAAGCACTAAACCTGTGGCCACCGAAAGATTGGGA 3957
Db 3883 TGATTCAGACGAGAGAGAAATTAAGCACTAAACCTGTGGCCACCGAAAGATTGGGA 3942
Qy 3958 CGGTGGAGCTCAATTTCCAGAGAGAGCGACAGACCTGTGACCGGAGATGTCAGCTA 4017
Db 3943 CTGTGGAGCTCAATTTCCAGAGAGAGCGACAGACCTGTGACCGGAGATGTCAGCTA 4002
Qy 4018 TGGAGAGCTTACCTGAGCTGTGTGCAAGATAGAGAGCTGTACCTGACGGTCCCATTT 4077
Db 4003 TGGAGAGCTTACCTGAGCTGTGTGCAAGAGAGAGAGAGCTGTACCTGAGCTTATTT 4062
Qy 4078 GGGCCAAATTTCTCACAAGATGAGACATTTCAACCCGTCTCTCTTATGAGCGCTTTG 4137
Db 4063 GGGCCAAATTTCTCACAAGATGAGACATTTCAACCCGTCTCTCTCTATGAGCGCTTTG 4122
Qy 4138 GACTCAAGAACCGGCTCTCAGATCTCATCAAAACAGCGCTGTCTCCGAAATCTTC 4197
Db 4123 GACTCAAGAACCGGCTCTCAGATCTCATCAAAACAGCGCTGTCTCCGAAATCTTC 4182
Qy 4198 CGGCGAGTTTTCAGTACAAAGTTTGTCTCATTTCACTTCAACCCATCTCCACAGAGAC 4257
Db 4183 CGGCGAGTTTTCAGTACAAAGTTTGTCTCATTTCACTTCAACCCATCTCCACAGAGAC 4242
Qy 4258 TGAAGTGAATTAATTAATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGAATCCGAG 4317
Db 4243 TGAAGTGAATTAATTAATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGAATCCGAG 4302
Qy 4318 TGAAGTGAATTAATTAATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGAATCCGAG 4377
Db 4303 TGAAGTGAATTAATTAATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGAATCCGAG 4362
Qy 4378 GACTTATATCTGAGCTCGCCCATTTGGACACCGCTTACCTTACCCGCTGTAAATTA 4437
Db 4363 GACTTATATCTGAGCTCGCCCATTTGGACACCGCTTACCTTACCCGCTGTAAATTA 4422
Qy 4438 GTGTAAATCAATAAACCAGTTGATTCGTTCAAGTTGAATTTGATCTCTCTCTCTTA 4497
Db 4423 GTGTAAATCAATAAACCAGTTGATTCGTTCAAGTTGAATTTGATCTCTCTCTTA 4482
Qy 4498 TCTTATC-GGTTACCATGTTTATAGCTTACATTAATCTGTTGGCTTCCGAGTA 4556
Db 4483 TCTTATC-GGTTACCATGTTTATAGCTTACATTAATCTGTTGGCTTCCGAGTA 4540
Qy 4557 AAAGACTTACGTATCGGAGTTTAAACCCCTAGTATGAGTTGACCACTCCCTCTGCGCG 4616
Db 4541 -----ATACCCCTAGTATGAGTTGACCACTCCCTCTATGCGCGC 4581
Qy 4617 TCGCTGCTCGTGGTGGGCGCTGCGAGCAAAAGTCCGAGAGCGAGAGCTCTCTGCGC 4676
Db 4582 TCGCTGCTCGTGGTGGGCGCGCGAGAGAGAGCTCTGCGAGAGCTCTCTGCGCA 4641
Qy 4677 GGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4718
Db 4642 GGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4683

```

```

RESULT 3
AAF23748
ID AAF23748 standard; DNA; 4722 BP.
XX AC AAF23748;
XX

```

```

DT 28-MAR-2001 (first entry)
XX
DE AAV3B DNA sequence.
XX
KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX atherosclerosis; sickle cell anemia; thalassemia;
XX blood clotting disorder; diabetes; ss.
XX
OS Adeno associated virus.
XX
PN US6156303-A.
XX
PD 05-DEC-2000.
XX
PF 11-JUN-1997; 97US-0873168.
XX
PR 11-JUN-1997; 97US-0873168.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Russell DW, Rutledge EA;
XX
DR WPI; 2001-060164/07.
XX
PT Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX
PS Example 2; Fig 1; 50bp; English.
XX
CC The present invention relates to adeno-associated virus serotypes. The
XX present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX
SQ Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;

Query Match 67.9%; Score 3205; DB 22; Length 4722;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 3872; Conservative 0; Mismatches 820; Indels 55; Gaps 12;

Qy 2 TGGCCACTCCCTCTCTGCGCGCTGCTCGCTCGTGGGCGCTGCGAGCAAAAGTCCGA 61
Db 1 TGGCCACTCCCTCTATGCGCACTGCTCGCTGCTGGGCGCTGCGAGCAAAAGTCCGA 60
Qy 62 GAGGAGAGCTCTGCTGCTGCGCGCGCCACCGAGAGAGAGCGCGCGAGAGAGAGTGG 121
Db 61 GAGGAGAGCTGCTTGGACGCTCGCGCCCAAGAGAGAGAGAGTGGCGCAAGAGAGTGG 120
Qy 122 GCAACTCATCTAGAGGGGTAATCGGAGAGCGGCTCCAGCGCTGCGCGAGCGCTGAC 181
Db 121 CCAACTCATCTAGAGGTAATGAGAGAGTACGTAACGCAAGCGCGCGAGAGAGCAC 180
Qy 182 GTAATTAAGCTATAGAGGAGTGTCTGTATTAAGTGTACAGTGAAGTCTTTTGAGCA 241
Db 181 GCCTACCA-----GCTGCTGAGAGTCAAGTGA-CCCTTTTGGAGCA 222
Qy 242 TTTTGGAGACCAACGTCGCTATTTAGGGTATATATGCGGAGTGAAGC-AGCAGATCTC 300
Db 223 GTTTGGAGACCAACGTCGCGCGCTGAGGGTATATCTGAGAGTGAACGAGAGAGCTC 282
Qy 301 CATTTGACCGGAAATTTGAACGAGCAGAGCATGCGCGGCTTTACAGATCGTGAT 360
Db 283 CATTTGACCGGAAATTTGAACGAGCAGAGCATGCGCGGCTTTACAGATGTCTCT 342
Qy 361 CAAGGTCCGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 343 GAAGGTCCGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402

```

QY 421 GGTGGCCGAGAGAGATGGAGCTGCCCGGATTTGACATGATCTGATTTGA 480
DB 403 GGTGGCCGAGAGAGATGGAGCTGCCCGGATTTGACATGATCTGATTTGA 462
QY 481 GCAGGACCCCTGACCGTGGCCGAGAACTGACGCGACTTCTTGCTCAATGGGCG 540
DB 463 GCAGGACCCCTGACCGTGGCCGAGAACTTCAAGCGAGTTCTTGCTGAGTGGCCCG 522
QY 541 CGTGAATGAGGCCCCGAGGCCCCCTTTCTTTGTTCACTTGAAGAGGAGTCTTACTT 600
DB 523 CGTGAATGAGGCCCCGAGGCCCCCTTTCTTTGTTCACTTGAAGAGGAGTCTTACTT 582
QY 601 CCACCTCCATATTTCTGGTGAAGACGAGGGGATCAAAATCAATGAGTGGGCGCTTCTCT 660
DB 583 CCACCTGACGCTGCTGATTTGAACCATGAGGATCAAAATCAATGAGTGGGCGCTTCTCT 642
QY 661 GAGTCAGATTAGGAGCAAGCTGTGACAGACCATCTACCGCGAGTCCGACCGACTTCC 720
DB 643 GAGTCAGATTAGGAGCAAGCTGTGACAGACCATCTACCGCGAGTCCGACCGACTTCC 702
QY 721 CAATGCTTCCGCGTGAACCAAGACGCTTAATGCGCGGAGACGAGGAACTGTTGGA 780
DB 703 GAATCTGTTCCGCGTGAACCAAAACGCGAAATGCGCGGAGACGAGGAACTGTTGGA 762
QY 781 CGAGTGTACATCCCGCACTACTCTGCGCCAGACTGAGCGAGTCCGAGTGGGCGTG 840
DB 763 CGACTGTACATCCCGCACTACTCTGCGCCAGACTGAGCGAGTCCGAGTGGGCGTG 822
QY 841 GACTTAACATGAGAGATATATAGCGCTGTTTGAACCTGAGCGCAACCGGCTCTG 900
DB 823 GACTTAACATGAGAGATATATAGCGCTGTTTGAACCTGAGCGCAACCGGCTCTG 882
QY 901 GGGCGACACCTGACCCAGCTGACGCGACCAAGAGAGAACTGGAACCC 960
DB 883 GGGCGACACCTGACCCAGCTGACGCGACCAAGAGAGAACTGGAACCC 942
QY 961 CAATTTGACGCGCGCTGATCCGCTCAAAAACCTCCGCGCTCACTAGAGTGGTCTCG 1020
DB 943 CAATTTGACGCGCGCTGATCCGCTCAAAAACCTCCGCGCTCACTAGAGTGGTCTCG 1002
QY 1021 GTGCTGTGTGACCGGCGCATCACCTCGAGAGCAATGATTCAGAGAGACGAGGCTC 1080
DB 1003 GTGCTGTGTGACCGGCGCATCACCTCGAGAGCAATGATTCAGAGAGACGAGGCTC 1062
QY 1081 GTATATCTCTTCAACCGCGCTTCCAACTGCGCGTCCCAATGCAAGCGCTCTGACAA 1140
DB 1063 GTATATCTCTTCAACCGCGCTTCCAACTGCGCGTCCCAATGCAAGCGCTCTGACAA 1122
QY 1141 TGCCGAGCAAGATCAATGCGCTGACCAAAATCGCGCGCTCACTGATGAGCCCGCTCC 1200
DB 1123 TGCCGAGCAAGATCAATGCGCTGACCAAAATCGCGCGCTCACTGATGAGCCCGCTCC 1182
QY 1201 GCCCGAGCAATTAACCAACCGCATCTACCGCATCTGAGTGAAGCTTCACTGACAA 1260
DB 1183 GCCCGAGCAATTAACCAACCGCATCTACCGCATCTGAGTGAAGCTTCACTGACAA 1242
QY 1261 TGCCGAGCAAGATCAATGCGCTTCTCGCTGAGGCGCAAAAAGTTCGAGAAAGCAAC 1320
DB 1243 GCAATGAGCGGCGCTCTGCTTCTGAGGCTGAGGCGCAAAAAGTTCGAGAAAGCAAC 1302
QY 1321 CATCTGCTGTTTGGGCGGACCAACGAGCAAGCAACCAATCGCGAAAGCATCGGCA 1380
DB 1303 CATCTGCTTCTTGGGCGGACCAACGAGGAAACCAACATCGCGAAAGCATCGGCA 1362
QY 1381 CGCGTGCCTTCTTACGCGCTGCTCAACTGAGCAACATGAGAACTTTCTCTTCAATGAT 1440
DB 1363 CGCGTGCCTTCTTACGCGCTGCTCAACTGAGCAACATGAGAACTTTCTCTTCAATGAT 1422
QY 1441 CGTGAACAAATGATCTGCTGAGAGAGGAGCAATGACGCGCAAGTCTGAGATC 1500
DB 1423 CGTGAACAAATGATCTGCTGAGAGAGGAGCAATGACGCGCAAGTCTGAGATC 1482

QY 1501 CGCCAAAGCCATTTCTCGGCGGACAGAGGTCGCGTGAACCAAAAGTGCAGTCTCGC 1560
DB 1483 CGCCAAAGCCATTTCTCGGCGGAGCAAGGTCGCGTGAACCAAAAGTGCAGTCTCGC 1542
QY 1561 CGAATGACCCCAACCCCGTGAATGCTACCTTCAACACCAACATGTCGCGGTGATGA 1620
DB 1543 CGAATGACCCCAACCCCGTGAATGCTACCTTCAACACCAACATGTCGCGGTGATGA 1602
QY 1621 CGGAAACAGACACCTTGAAGACAGAGCGCGTTCAGAGACCGGATGTTCAATTTGA 1680
DB 1603 CGGAAACAGACACCTTGAAGACAGAGCGCGTTCAGAGACCGGATGTTCAATTTGA 1662
QY 1681 ACTCACCCGCGCTGAGAGATGATTTGGCAAGTGAACAAACAGAGATCAAGATTTGA 1740
DB 1663 ACTCACCCGCGCTTGAAGATGATTTGGCAAGTGAACAAACAGAGATCAAGATTTGA 1722
QY 1741 CTTCGCTGGGCGCAAGATCACTGACCGAGTGGCGCATGATTTCTACCTCAAAAAGG 1800
DB 1723 CTTCGCTGGGCGCTTCGATCACTGACCTGACGTCGCTCATGATTTCTACCTCAAAAAGG 1782
QY 1801 TGGAGCAACAAAGAACCCGCGCGATGACGCGGATTAAGAGGAGCCCAAGGCGGCTG 1860
DB 1783 TGGAGCAACAAAGAACCCGCGCTTCATGACGCGATGTAAGGAGCAAAACGACATG 1842
QY 1861 CCCCTGATCGCGGATTCATGACGCTCAGACGCGAAGAGCTCCGATGACTTTGCGCA 1920
DB 1843 CAGTCACTTGGCGACCCGCAACGCTCAGACGCGGAG---AGCACCGCGGAGCTACCGGAG 1899
QY 1921 CAGGTACAAACCAATGTTCTGCTGACCGCGGATGCTTCAATGCTGTTTCCCTGCA 1980
DB 1900 CAGGTACAAACCAATGTTCTGCTGACCGCGGATGATGATGATCTTTTCCCTGTA 1959
QY 1981 GACATGAGAGAGATGATGATGATTTCAACATTTGCTCAGACGAGGAGCAGAGACTG 2040
DB 1960 GACATGAGAGAGATGATGATGATTTCAACATTTGCTCAGAGTGTGATTAAGAGACTG 2019
QY 2041 TTGAGAGCTTCCCGCGCTGTCAGATCTCAAC-----CGGTGCTCAGAAAGAGAC 2094
DB 2020 TTGAGAGATGCTTCCGAGATGTCAGATCTCAACCCGTTCTGCTGTCAGAAAGAGAC 2079
QY 2095 GTATCGAAACCTGTCGCAATTTATCATCTGCTGGGCGGAGCTCCGAGATGCTGCTC 2154
DB 2080 GTATCGAAACCTGTCGCAATTTATCATCTGCTGGGAGAGGACCCGAGATGCTGCTC 2139
QY 2155 GGCCTGATCTGATCAAGTGAACCGGATGACGCTGCTTCAAGCAATTAATGACTTA 2214
DB 2140 GGCCTGATTTGGCAATGATGACTGATGACTGCTGCTTCAAGCAATTAATGACTTA 2199
QY 2215 AACAGATAGCTGCGAGTGTATCTTCAAGATTTGCTGAGAGCAACCTCTGAGG 2274
DB 2200 AACAGATAGCTGCTGAGAGTGTATCTTCAAGATTTGCTGAGAGCAACCTTCTGAGG 2259
QY 2275 GCATTTGCGAGTGTGAGACTTGAACCTGAGACCCCGAAGCCCAAGCCCAACAGCAAA 2334
DB 2260 GCATTTGCTGATGCTGAGCTTGAACCTGAGATCTCTCAACCAAGCCCAACCAAC 2319
QY 2335 AGCAGACAGCGCGCGGAGTCTGCTGCTTCAAGATCTGAGACCTTCAACG 2394
DB 2320 AGCAGACAGCGCTGAGGCTTGTGCTTCCGAGTTTCAAAATCTCTGAGACCGGTAAG 2379
QY 2395 GACTGCAACAGGAGAGCCCGTCAACGCGCGAGCAAGCGGCTTCAAGCAAGCAAG 2454
DB 2380 GACTGCAACAGAGAGCCCGTCAACGAGGCGAGCGGACGCTTCAAGCAAGCAAG 2439
QY 2455 CTTACGACAGAGCTTCAAGCGGAGTGAACATCTGATCTGCGGTTATACAGCCGAGCG 2514
DB 2440 CTTACGACAGAGCTTCAAGCGGAGTGAACATCTGATCTGCGGTTATACAGCCGAGCG 2499
QY 2515 CGAGTTTCAAGAGCGCTTCAAGAGATGATGCTTTTGGGAGCACTTGGCGAGAGCG 2574
DB 2500 CGAGTTTCAAGAGCGCTTCAAGAGATGATGCTTTTGGGAGCACTTGGCGAGAGCG 2559
QY 2575 TCTTCCAGGCAAGAGCGGCTTCTGAACTCTCGATCTGCTTGAAGAGGCGCTAAG 2634

Db	2560	TCCTCCAGGCGCAAAAAGAGATCTTGAGCCTCTTGCTGTGATGAGAAAGACAGCTAA	2619
Qy	2635	CGGCTCTCGAANAANAAGTCGCTAGAGATGCGCAAGAGCCAGATCTCTCTCGG	2694
Db	2620	CGGCTCTCGAANAANAAGCTGTAGATCACTCTCTAGAGAAACCGGACTCACTCTG	2679
Qy	2695	GCATCGGCAAGACAGGCGCAGAGCCCGCTAAAAAGAGACTCAATTTTGTCTAGCTGGG	2754
Db	2680	GTTGTGCAAAATCGGCGAAACAGCTCTGCGAANAACCTAAATTTGGCTCAGCTGGG	2739
Qy	2755	ACTCAGAGTCACTGCTCCGATCCACACCTCTCGAGAACCTCCGACACCCCGCTGCTG	2814
Db	2740	ACTCAGAGTCACTGCTCCGATCCACACCTCTCGAGAACCAACGACGACCCCAAGTT	2799
Qy	2815	TGGGACCTACTACAAATGGCTTCAAGCGGTGCGGACCAATGCGAGACAAATACAAAGCG	2874
Db	2800	TGGGATCTAAATACATAGCTTCAAGCGGTGCGGACCAATGCGAGACAAATACAAAGCG	2859
Qy	2875	CCGACGAGTGGGATATGCTCAGAGAAATTTGGCAATTCGATTCACATGGCTGGGCGCA	2934
Db	2860	CCGATGAGATGGGATATTCCTCAGAGAAATTTGGCAATTCGATTCGATTCGATTCGATTC	2919
Qy	2935	GAGTCATCACACCAAGACCCCGACCTGGGCTTTGCCACCTACATATACACCTCTACA	2994
Db	2920	GAGTCATCACACCAAGACCCCGACCTGGGCTTTGCCACCTACATATACACCTCTACA	2979
Qy	2995	AGCAAAATCTCAGCTTCAACGAGGCGGCGACAGACCAACCACTACTTGGCTACAGCA	3054
Db	2980	AGCAAAATCTCAGCTTCAACGAGGCGGCGACAGACCAACCACTACTTGGCTACAGCA	3036
Qy	3055	CCCCCTGGGGATATTTGATTTTCAACAGATTTCACTGCTCACTTTTCAACAGTACGCTG	3114
Db	3037	CCCCCTGGGGATATTTGATTTTCAACAGATTTCACTGCTCACTTTTCAACAGTACGCTG	3096
Qy	3115	AGGACTCATCAACCAAAATTTGGGATTTCCGCGCAAGAGACTCACTCAAACTCTTCA	3174
Db	3097	AGGACTCATCAACCAAAATTTGGGATTTCCGCGCAAGAGACTCACTCAAACTCTTCA	3156
Qy	3175	ACATCCAGTCAGAGAGATGACCAAGATATGAGCGCTCAACCACTATGCTAAATACCTTA	3234
Db	3157	ACATCCAGTCAGAGAGATGACCAAGATATGAGCGCTCAACCACTATGCTAAATACCTTA	3216
Qy	3235	CCAGCAGGCTCAAGCTTCTCGGACTCGAGATACAGCTTCGATCGCTCGGCTCTG	3294
Db	3217	CCAGCAGGCTCAAGCTTCTCGGACTCGAGATACAGCTTCGATCGCTCGGCTCTG	3276
Qy	3295	CGACACGAGGCTGCTCTCCCTCCGCTTCCGCGGAGAGCTTTCATGATTCGCAATACGCT	3354
Db	3277	CGACACGAGGCTGCTCTCCGCTTCCGCGGAGAGCTTTCATGATTCGCAATACGCT	3336
Qy	3355	ACCTGAGCTCAACATATGAGCAGCAGACCGCTGGAGCTTCACTCTTTTACTGCTGGAAT	3414
Db	3337	ACCTGAGCTCAACATATGAGCAGCAGACCGCTGGAGCTTCACTCTTTTACTGCTGGAAT	3396
Qy	3415	ATTTCCCTCTCAGATGCTGAGAACGCGCAACACTTACCTTCAGCTACACCTTTAGG	3474
Db	3397	ATTTCCCTCTCAGATGCTGAGAACGCGCAACACTTACCTTCAGCTACACCTTTAGG	3456
Qy	3475	AACTGCTTTTCAACAGAGCTACGCGCAGACAGCAGACCTGGAACCGGCTGATGATCTC	3534
Db	3457	AACTGCTTTTCAACAGAGCTACGCGCAGACAGCAGACCTGGAACCGGCTGATGATCTC	3516
Qy	3535	TCATCGACCAATATCTGTATTTACTGTAACGAACTCA--AAATCAGTCCGGAAGTGGC	3591
Db	3517	TCATCGACCAATATCTGTATTTACTGTAACGAACTCA--AAATCAGTCCGGAAGTGGC	3576
Qy	3592	AAAAAAGAACTGCTGTTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3651
Db	3577	AAAAAAGAACTGCTGTTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3636
Qy	3652	ACTGGCTACTGAGCCCTGTTATCGAGAGCGCGCTTTTAAACAAAAACAGACACA	3711
Db	3637	ACTGGCTACTGAGCCCTGTTATCGAGAGCGCGCTTTTAAACAACTGTAAACAGACACA	3696
Qy	3712	ACAAAGCAATTTTACTGAGATGCTGCTTCAAAATATATACCTCATATGGCGGATCCCA	3771
Db	3697	ACAAAGCAATTTTACTGAGATGCTGCTTCAAAATATATACCTCATATGGCGGATCCCA	3756
Qy	3772	TCATCAACCTGCGACCTGCTATGAGCTTCAACAAAGCAGCAAGAACTTCTTCCCA	3831
Db	3757	TCATCAACCTGCGACCTGCTATGAGCTTCAACAAAGCAGCAAGAACTTCTTCCCA	3816
Qy	3832	TGACCGGTCTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAAACTGTACATTGACA	3891
Db	3817	TGACCGGTCTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAAACTGTACATTGACA	3876
Qy	3892	ATGTCATGATTAAGAGAGAAATTAACCCCTAACCTGTCGACCCGGAAGAT	3951
Db	3877	ATGTCATGATTAAGAGAGAAATTAACCCCTAACCTGTCGACCCGGAAGAT	3936
Qy	3952	TTGGGACCGTGGGAGCAATTTTCAGAGCAGCAGACACCTGCGACCGGAGATGTGC	4011
Db	3937	TTGGGACCGTGGGAGCAATTTTCAGAGCAGCAGACACCTGCGACCGGAGATGTGC	3996
Qy	4012	ATGCTATGAGGACATTAACCTGATGATGATGAGGAGATGAGAGCTGACCTGACAGGTC	4071
Db	3997	ATGCTATGAGGAGCTTAACCTGATGATGATGAGGAGATGAGAGCTGACCTGACAGGTC	4056
Qy	4072	CCATTTGGGCAAAATTTCTTCAACAGATGAGACATTTTACCCGCTCTCTTATGGCG	4131
Db	4057	CCATTTGGGCAAAATTTCTTCAACAGATGAGACATTTTACCCGCTCTCTTATGGCG	4116
Qy	4132	GCTTTGAGCTCAAGAACCCGCTCTCAAGATCTCAATCAAAACAGCGCTGCTGCGCA	4191
Db	4117	GCTTTGAGCTCAAGAACCCGCTCTCAAGATCTCAATCAAAACAGCGCTGCTGCGCA	4176
Qy	4192	ATCTTCGCGGAGTTTACGCTAACAAAGTTTCTTCAATTCATACCAATCTGACAG	4251
Db	4177	ATCTTCGCGGAGTTTACGCTAACAAAGTTTCTTCAATTCATACCAATCTGACAG	4236
Qy	4252	GACAAGTGTGTGAAATTTGATTTGGAGCTGCGAAGAAACACCAAGCGCTGGAATC	4311
Db	4237	GACAAGTGTGTGAAATTTGATTTGGAGCTGCGAAGAAACACCAAGCGCTGGAATC	4296
Qy	4312	CCGAGTGCAGTACATCCAAATTTATGCAAAATCTGCAAGCTTGAATTTTACTGAGACA	4371
Db	4297	CCGAGTGCAGTACATCCAAATTTATGCAAAATCTGCAAGCTTGAATTTTACTGAGACA	4356
Qy	4372	ACAATGACCTTTATGAGCTTCCGCTGCGCAATTTGCAACCGGTTACCTTACCCGCTGT	4431
Db	4357	ACAATGACCTTTATGAGCTTCCGCTGCGCAATTTGCAACCGGTTACCTTACCCGCTGT	4416
Qy	4432	AATTTAGCTTATCAATTAACCGGTTGATTTGCTTCACTTGAACCTTTGG--TTCCTGTC	4490
Db	4417	AATTTAGCTTATCAATTAACCGGTTGATTTGCTTCACTTGAACCTTTGG--TTCCTGTC	4475
Qy	4491	CTTCTATCTTATC--GATTAACATGATTTAGCTTACATTAAGT-----CTTGGTT	4543
Db	4476	CTTCTATCTTATC--GATTAACATGATTTAGCTTACATTAAGT-----CTTGGTT	4535
Qy	4544	GCGCTTCGCGATTAAGAAC-----TTACGCTATCGGTTTACCTTATGATGAG	4591
Db	4536	GCGCTTCGCGATTAAGAAC-----TTACGCTATCGGTTTACCTTATGATGAG	4576
Qy	4592	AGTTTCCCACTCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4651
Db	4596	AGTTTCCCACTCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4655
Qy	4652	GACAGCGGAGAGCTGCTCTGCGGCGCCACAGAGCGAGCGAGCGAGCGAGCGAG	4711
Db	4656	TCCGTCTGCGAGCTTTGCTGCGAGGCGCCACAGCGAGCGAGCGAGCGAGCGAG	4715
Qy	4712	TGGGCAA 4718	
Db	4716	TGGGCAA 4722	

QY 717 TGCCCACTGGTTGCGGGTGAACAAGACGGGTATGCGCGGAGGGGGAACAAGTGG 776
DB 703 TGCCAAACTGGTTGCGGGTGAACAAGACGGGTATGCGCGGAGGGGGAACAAGTGG 762
QY 777 TGAAGAGTGTCTATCTCCCACTACTCTCTGCGCAAGTCTCAAGCCGAGCTGCACTGG 836
DB 763 TGAATGAGTGTCTATCTCCCACTACTCTCTGCGCAAGTCTCAAGCCGAGCTGCACTGG 822
QY 837 CGTGAATCAATGAGAGTATATATAGCGCTGTTGAACCTGCGCAAGCGCAACCGC 896
DB 823 CGTGAATCAATGAGAGTATATATAGCGCTGTTGAACCTGCGCAAGCGCAACCGC 882
QY 897 TCGTGGCGAGCACTGACCGCAAGTCAAGCCGAGCAAGCAAGCAAGCAAGTGA 956
DB 883 TCGTGGCGAGCACTGACCGCAAGTCAAGCCGAGCAAGCAAGCAAGCAAGTGA 942
QY 957 ACCCAATTCTGACGCGCTGTATCTGCTCAAAAACCTCGCGCTCATGAGTGG 1016
DB 943 ATCCCAATCTGATGCGCGGTATCATCAAAAACCTCGCGCTCATGAGTGG 1002
QY 1017 TCGGGTGTGTGTGAGCCGGGATCATCTCCGAGAGAGTGTATCAAGAGACCAAG 1076
DB 1003 TCGGGTGTGTGTGAGCCGGGATCATCTCCGAGAGAGTGTATCAAGAGACCAAG 1062
QY 1077 CCTGTATCATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCGCTCTGG 1136
DB 1063 CCTGTATCATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCGCTCTGG 1122
QY 1137 ACAATGCGGCAAGATCATGAGCGCTGAACAATCGCGCGCACTACTGTTAGCGCCG 1196
DB 1123 ACAATGCGGCAAGATCATGAGCGCTGAACAATCGCGCGCACTACTGTTAGCGCCG 1182
QY 1197 CTCCGCGCGGCAATTTAAACCAAGCATCTACCGCATCTGAGCTCAAGCGCTAG 1256
DB 1183 AGCCCGTGAAGCAATTTCAAGCATCTGAGATTTAAATTTTGAACCTAAAGCGGTAG 1242
QY 1257 AACCTGCTACGCGCGCTCTGCTCTGCTGCGGTGGGCGAGAAAAGTTGCGGAGCGCA 1316
DB 1243 ATCCCAATATGAGGCTTCTGCTCTGCTGCGGTGGGCGAGAAAAGTTGCGGAGCGCA 1302
QY 1317 ACACCACTGTGTGTTGGGCGCGCAACAAGCGCAAGCAACATCGCGGAGCGCACTG 1376
DB 1303 ACACCACTGTGTGTTGGGCGCGCAACAAGCGCAAGCAACATCGCGGAGCGCACTG 1362
QY 1377 CCAAGCGCGCTCTTCTAGCGCTGCTCACTGAGCAATGAGAACTTCTCTCAATG 1436
DB 1363 CCAAGCGCGCTCTTCTAGCGCTGCTCACTGAGCAATGAGAACTTCTCTCAATG 1422
QY 1437 ATTGCGTCAAGAGTGTATCTGAGGAGGAGGCAAGATGACGCGCAAGTGTGG 1496
DB 1423 ACTGTGTCAAGAGTGTATCTGAGGAGGAGGAGGAGATGACGCGCAAGTGTGG 1482
QY 1497 AGTCCGCAAGGCACTTCTGCGGCGCAAGGAGCGCTGAGCAACAAGTGAAGTCTG 1556
DB 1483 AGTCCGCAAGGCACTTCTGCGGCGCAAGGAGCGCTGAGCAACAAGTGAAGTCTG 1542
QY 1557 CCGCCCAAGTCAAGCCCAAGCGCGCTGATCTGCTCAACCAACATGTGGCGCTGA 1616
DB 1543 CCGCCCAAGTCAAGCCCAAGCGCGCTGATCTGCTCAACCAACATGTGGCGCTGA 1602
QY 1617 TTGACGGGAACAGACCACTTGAAGCAGAGCGCTGAGAGCCGATGTTCAAT 1676
DB 1603 TTGACGGGAACAGACCACTTGAAGCAGAGCGCTTGAAGCAGAGCGATGTTCAAT 1662
QY 1677 TTGAACTCAAGCGCGCTGTGAGCATGACTTTGCAAGGAGTGAACAAGCAAGTCAAG 1736
DB 1663 TTGAACTCAAGCGCGCTGTGAGCATGACTTTGCAAGGAGTGAACAAGCAAGTCAAG 1722
QY 1737 AGTTCTTCCGCTGGGCGAGATCAAGTGAACGAGGAGCGCATGAGTTTAACTCAAGAA 1796
DB 1723 ACTTTTCCGCTGGGCGAGATCAAGTGAACGAGGAGCGCATGAGTTTAACTCAAGAA 1782

QY 1797 AGGTTGAGCCCAACAAGAACCCCGCCGATGACCGCGATTAAGCGAGCCCAAGCGG 1856
DB 1783 AGGTTGAGCCCAACAAGAACCCCGCCGATGACCGCGATTAAGCGAGCCCAAGCGG 1842
QY 1857 CCGTCCCTCAAGTCCGCGATTCATGAGGTCAAGCGCGAAGAGTCTCGGTGCACTTGG 1916
DB 1843 TCGCGAGTCAAGTGGCGAGCATCAAGCTCAAGCCCGAAG--AGCTGATCACTAGC 1899
QY 1917 CCGAGAGTCAACAAGTGTCTGTCAGCGCGGCAAGTCTGAGTGTCTGCTTCCCT 1976
DB 1900 CAGACAGTCAACAAGTGTCTGTCAGCGCGGCAAGTCTGAGTGTCTGCTTCCCT 1959
QY 1977 GCAAGCATCGAGAGAAATGAATCAAGATTTCAACTTGTCTTCAAGCAAGGAGCAGAG 2036
DB 1960 GCAAGCATCGAGAGAAATGAATCAAGATTTCAACTTGTCTTCAAGCAAGGAGCAGAG 2019
QY 2037 ACTGTTCAAGTGTCTTCCCGCGGTCTCAAGTCTCAAGCGGT--CGTCAAGAGAGA 2093
DB 2020 ACTGTTCAAGTGTCTTCC--CGTCAAGAGTCTCAAGCGGTCTGTCGCAAAAAGG 2076
QY 2094 CGTATGAGAACTGTGTGATCATCTGCTGCGGCGGCGCTCCGAGATTTGCTGCT 2153
DB 2077 CGTATGAGAACTGTGTGATCATCTGCTGCGGCGGCGCTCCGAGATTTGCTGCT 2133
QY 2154 CCGCTGCGATCTGTCAAGTGAACCTGAGTCACTGTGTTCTGAGCAATTAATGACTT 2213
DB 2134 CTGCTGCGATCTGTCAAGTGAATTTGAGTGAATGATGATCTTGAACATTAATGATTT 2193
QY 2214 AAACAGATATGCTGCGCATGTTATCTTCAAGATTTGCTGAGAGCAACTCTGAG 2273
DB 2194 AAATCAAGTATGCTGCGCATGTTATCTTCAAGATTTGCTGAGAGCAACTCTGAG 2253
QY 2274 GCGATTCGAGAGTGTGAGCACTTGAACCTGAGACCCCGCAAGCCCAAGCAAGCA 2333
DB 2254 GCGATTCGAGAGTGTGAGCACTTGAACCTGAGACCCCGCAAGCCCGAGCGG 2313
QY 2334 AAGCAGAGCAGCGCGCGGTCTGATCTTCTGCTGCTAAGATCTCGGACCTTCAAC 2393
DB 2314 CATAAGAGCAGCAGCGGCTTGTGCTTCTCGGCTAAGATCTCGGACCTTCAAC 2373
QY 2394 GGACTGACAGAGGAGGAGCCCGTCAAGCGCGGAGCGAGCGGCTCTGAGCAGCAAG 2453
DB 2374 GGACTGACAGAGGAGGAGCCCGTCAAGCGGAGCGAGCGGCTCTGAGCAGCAAG 2433
QY 2454 GCTTACAGCAGAGTCAAGCGGCTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2513
DB 2434 GCTTACAGCAGAGTCAAGCGGCTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2493
QY 2514 GCGAGTTCAGGAGCGCTGCAAGAGATACGCTTTTGGGGGCAACCTCGGGCAGAGA 2573
DB 2494 GCGAGTTCAGGAGCGCTTAAAGAGATACGCTTTTGGGGGCAACCTCGGGCAGAGA 2553
QY 2574 GTCTTCAGGCGCAAGAGCGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTG 2633
DB 2554 GTCTTCAGGCGCAAGAGCGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTG 2613
QY 2634 ACGGCTCTGAGAAAGAAAGTCCGCTGAGAGCACTGCTGCTGCTGCTGCTGCTGCTG 2693
DB 2614 ACGGCTCTGAGAAAGAAAGTCCGCTGAGAGCACTGCTGCTGCTGCTGCTGCTGCTG 2673
QY 2694 GCGATCGAGAGCAGAGCGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2753
DB 2674 GCGATCGAGAGCAGAGCGGCTTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2733
QY 2754 GACTCAGTCAAGTCCCGATTCACACTCTGCTGAGAACTTCAAGCAACCCCGCTGCT 2813
DB 2734 GACTCAGTCAAGTCCCGATTCACACTCTGCTGAGAACTTCAAGCAACCCCGCTGCT 2793
QY 2814 GTGGGACCTTAATCAATGCTTCAAGCGGCTGCGCAGCAATGCGAGCAATTAAGAGG 2873
DB 2794 GTGGGACCTTAATCAATGCTTCAAGCGGCTGCGCAGCAATGCGAGCAATTAAGAGG 2853
QY 2874 GCCGACGAGTGGGTATGCTCAGGAATTTGGCATTTCAATGCTGAGCGGCGAC 2933

Db 2854 GCCACGAGAGTGGTAATTCCTCGGGAATTTGGCATTCGATTCACATGGATGGCGAC 2913
 Qy 2934 AGAGTCATCACACACGACCGGCACTTGGGCTTTGCCCATTAACATTAACACCTTCTAC 2993
 Db 2914 AGAGTCATCACACACGACCGGCACTTGGGCTTTGCCCATTAACATTAACACCTTCTAC 2973
 Qy 2994 AAGGAATTCGACGCTTCAACGGGGGGCGACCAACGACCACTTCTGGGCTTACAGC 3053
 Db 2974 AAGGAATTCGACGCTTCAACGGGGGGCGACCAACGACCACTTCTGGGCTTACAGC 3030
 Qy 3054 ACCCTGGGGGTAATTTGATTTCAACAGATTCACCTGCACTTTTCAACAGCTGCTG 3113
 Db 3031 ACCCTGGGGGTAATTTGATTTCAACAGATTCACCTGCACTTTTCAACAGCTGCTG 3090
 Qy 3114 CAGGACTATCAACACAAATTTGGGATTTCCGGCCCAAGAGACTTCAACTTCAACTCTTC 3173
 Db 3091 CAAGAGACTATCAACACAACTGGGGATTTCCGACCCCAAGAGACTTCAACTTCAACTCTTC 3150
 Qy 3174 AACATCAAGTCAGAGGGGTACAGACGAAATGATGGCGTCAACACATGGCTAATACCTT 3233
 Db 3151 AACATCAAGTCAGAGGGGTACAGACGAAATGATGGCGTCAACACATGGCTAATACCTT 3210
 Qy 3234 ACCGACGCTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTACGTCCTCGGCTCT 3293
 Db 3211 ACCGACGCTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTACGTCCTCGGCTCT 3270
 Qy 3294 GCGGACACGAGGCTGCTTCCCTCGGCTTCCGGCGGACGCTTCAATGATTTCCGCAATAGGC 3353
 Db 3271 GCGGATCAAGGATGCTTCCCTCGGCTTCCGGCGGACGCTTCAATGATTTCCGCAATAGGC 3330
 Qy 3354 TACCTGACGCTCAACATGAGCGACCAAGCGGAGGAGTTCATCTTTTACCTGCTGAA 3413
 Db 3331 TACCTGACGCTCAACATGAGCGAGTCAAGGAGTTCATCTTTTACCTGCTGAG 3390
 Qy 3414 TATTTCCCTTCTAGATGCTGAGAAACGGGCAACAACTTACCTTCACTGACCTTTGAG 3473
 Db 3391 TATTTCCCTTCTAGATGCTGAGAAACGGGCAACAACTTACCTTCACTGACCTTTGAG 3450
 Qy 3474 GAAATGCTTTCACAGAGCTACGCGGACGACGACGCTTGAACCGGCTGATGATTCCT 3533
 Db 3451 GAAATGCTTTCACAGAGCTACGCGGACGACGACGCTTGAACCGGCTGATGATTCCT 3510
 Qy 3534 CTATCGACCAATACCTGATTTACCTGAAACAACTCAAAATAGTCCGGAAGTCCCA 3593
 Db 3511 CTATCGACCAATACCTGATTTACCTGAAACAACTCAAAATAGTCCGGAAGTCCCA 3570
 Qy 3594 AACAGGACTTGTGTTTACGCGGTGAGTCTCAGCTGCAATGCTTGTTCAGCCCAAAAC 3653
 Db 3571 CAGTCAAGGCTTCAAGTTCCTCAGGCGGAGCGAGTCAATTCGCGAACCACTTAGAAC 3630
 Qy 3654 TGGTACTCTGACCTCTGTTATCGGACAGAGCGGCTTTTAAACAAACAGACACACAC 3713
 Db 3631 TGGTACTCTGACCTCTGTTATCGGACAGAGCGGCTTTTAAACAAACAGACACACAC 3690
 Qy 3714 AACGACATTTTACCTGAGTGTGCTTCAAAATTAACCTCAATGGGCGGATTCATC 3773
 Db 3691 AACGACATTTTACCTGAGTGTGCTTCAAAATTAACCTCAATGGGCGGATTCATC 3750
 Qy 3774 ATCAACCTTGCACTGCTATGAGCTCACAAAGACGACGAAACAAGTCTTTCCATG 3833
 Db 3751 GTGATCCGGGCTCGGACATGACCAAGACGACGAAAGTAAAGTTCCTCAG 3810
 Qy 3834 AGCGGTGTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAACATGCTGATTTGACAT 3893
 Db 3811 AGCGGTGTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAACATGCTGATTTGACAT 3870
 Qy 3894 GTCATGATTAACAGAGAAAGAAATTAAGCCATTAACCTTGGCCACGAAAGATTT 3953
 Db 3871 GTCATGATTAACAGAGAAAGAAATTAAGCCATTAACCTTGGCCACGAAAGATTT 3930
 Qy 3954 GGGACCGTGGACGTCATTTTCAAGACGACGACCAAGCCCTGCGACCGGATGTCAT 4013

Db 3931 GGTTCTGATCTACCAACTCTCAGAGAGGCAACAGACAAACAGACTACCGGATGTCAC 3990
 Qy 4014 GCTATGGAGACATTTACTGCAATGCTGTGCAAGATTAAGACCTGTAACCTGAGGCTCC 4073
 Db 3991 ACACAGGCGCTTCTTCCAGGATGCTGTGCAAGAGAGATGTAACCTTCAAGGCGCC 4050
 Qy 4074 AATTGGGCAAAATTCCTCAACAGATGACACTTTCACCCGCTCTCTTATGAGCGGC 4133
 Db 4051 ATCTGGCAAAAGATTCACACAGAGAGACATTTTCAACCCCTCTCCCTCATGGGTGA 4110
 Qy 4134 TTTGACTCAAGAACCGGCTCTCTCAGATCTCATCAAAAACAGCGCTTCTCTGCAAT 4193
 Db 4111 TTTGACTCAAGAACCGGCTCTCTCAGATCTCATCAAAAACAGCGCTTCTCTGCAAT 4170
 Qy 4194 CTTCGGGAGGATTTAGCTACAAAGTTGCTTCAATTCATCCCAATTTCTCCACAGA 4253
 Db 4171 CTTCGGGAGGATTTAGCTACAAAGTTGCTTCAATTCATCCCAATTTCTCCACAGA 4230
 Qy 4254 CAAGTGTGTGGAATTTGAATGGGAGCTGACAGAAAGAAACAGCAAGCGCTGGAATCCC 4313
 Db 4231 CAAGTGTGTGGAATTTGAATGGGAGCTGACAGAAAGAAACAGCAAGCGCTGGAATCCC 4290
 Qy 4314 GAATGCACTACACATTCATTAATGCAAAATTTGCCAAGCTTGAATTTACTGTGACAC 4373
 Db 4291 GAATGCACTACACATTCATTAATGCAAAATTTGCCAAGCTTGAATTTACTGTGACAC 4350
 Qy 4374 AATGCACTTATACAGCTGCTGCGCCATTTGGAGACCGCTTACCTTACCGCTCCGTGA 4433
 Db 4351 AATGCACTTATACAGCTGCTGCGCCATTTGGAGACCGCTTACCTTACCGCTCCGTGA 4410
 Qy 4434 TTAAGTGTATCAATTAACCGGTTGATTTGCTTCAAGTGAACCTTGTCTCTGCTT 4493
 Db 4411 TTAAGTGTATCAATTAACCGGTTGATTTGCTTCAAGTGAACCTTGTCTCTGCTT 4470
 Qy 4494 CTATCTTATGCTTACATGATTTATGCTTCAATTAACCTTGTCTCTGCTT 4553
 Db 4471 CTATCTTATGCTTACATGATTTATGCTTCAATTAACCTTGTCTCTGCTT 4530
 Qy 4554 ATAAAGACTTACGTATCGGGTTACCCCTAGATGAGATGAGCCCACTCCCTCTGCG 4613
 Db 4531 TACAGGA-----ACCCCTAGATGAGATGAGCCCACTCCCTCTGCG 4574
 Qy 4614 CGCTGCTGCTCGATGGGCTGCGGACCAAGGTTCCGCAACGGGCAAGCTTGTCT 4673
 Db 4575 CGCTGCTGCTCGATGGGCTGCGGACCAAGGTTCCGCAACGGGCTTGTGCGG 4634
 Qy 4674 GCGGCGCCCAACGAGGAGCGAGCGGCGGAGAGGGGAGGCA 4718
 Db 4635 GCGGCGCTCAGTGAAGGAGCGGCGGAGAGGGAGTGGCCA 4679

RESULT 5
 AAH26325
 ID AAH26325 standard; DNA; 8698 BP.
 XX
 AC
 AAH26325;
 XX
 DT
 02-OCT-2001 (first entry)
 XX
 DE
 Wild-type adeno-associated virus genome plasmid pAV2.
 XX
 KW
 AAV; vector; pAV2; packaging cell line; gene therapy; de.
 XX
 OS
 Adeno associated virus.
 XX
 PN
 MO200155361-A2.
 XX
 PD
 02-AUG-2001.
 XX
 PF
 26-JAN-2001; 2001MO-US02709.
 XX
 PR
 26-JAN-2000; 2000US-0178536.
 XX

1437 ATTCGTCGACAGATGATGATCTGTGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGG 1496
1423 ACTGTGTGACAAAGATGATGATCTGTGTGGAGAGAGGAGATGATACGGCCAAAGTCTGG 1482
1497 AGTCGCGCAAGGCAATCTCGGCGGAGCAAGGTGCGGTGACCAAAAGTCAAGTCTG 1556
1483 AGTCGCGCAAGGCAATCTCGGAGGAAAGCAAGGTGCGGTGACCAAAAGTCAAGTCTG 1542
1557 CCGGCGCAAGTGAACCCCGGCGGTGATGCTCACTCAACCAAGATGCGCGGTG 1616
1543 CCGGCGCAAGTGAACCCCGGCGGTGATGCTCACTCAACCAAGATGCGCGGTG 1602
1617 TTGACGGAACAGACACCACTTCGAGCAAGCCGTTGACAGACCGGATGTTCAAT 1676
1603 TTGACGGAACCTCAACCACTTCGAGCAAGCCGTTGACAGACCGGATGTTCAAT 1662
1677 TTGACCTCAACCCCGGCGGTGAGATGATCTTTGGCAAGGTGACAAAGAGGATCAAG 1736
1663 TTGAACTCAACCCCGGCGGTGATGATCTTTGGCAAGGTGACCAAGAGGATCAAG 1722
1737 AGTTCTTCGCTGGGCGGAGATCACTGACCGAGGTGCGCATGATGATCTTACGTCAGAA 1796
1723 ACTTTTTCGCGTGGGCAAAAGATCACTGTTGAGGTGAGCATGAAATTTACGTCAGAA 1782
1797 AGGCTGAGCAACAAAGACCCGCGCGGATGACGCGGATTAAGCGAGCCCAAGCGGG 1856
1783 AGGCTGAGCAACAAAGACCCGCGCGGATGACGCGGATTAAGGTAGGACCAAGCGGG 1842
1857 CTTGCGGCTCACTGCGGATCACTGACGTCAGACGCGGAAAGAGCTCGGTGACCTTGG 1916
1843 TGCCGAGTCACTGCGGATCACTGACGTCAGACGCGGAAAGAGCTCGGTGACCTTGG 1899
1917 CCGGAGGTACCAAAAGATGATCTGTCAGACGCGGATGATGATGATGATGATGATGATGAT 1976
1900 CAGACAGGTACCAAAAGATGATCTGTCAGACGCGGATGATGATGATGATGATGATGATGAT 1959
1977 GCAAGACATGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2036
1960 GCAAGACATGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2019
2037 ACTGTCAGAGTCTTCCCGGCGGTGACAGATCTCAACCGGT--CGTCAGAAAGAG 2093
2020 ACTGTCAGAGTCTTCCCGGCGGTGACAGATCTCAACCGGTCTGTCAGAAAGAG 2076
2094 CGTATCGGAACTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2153
2077 CGTATCGGAACTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2133
2154 CCGGCTGCGATCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2213
2134 CTGCTGCGATCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2193
2214 AAACCAAGTATGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2273
2194 AAACCAAGTATGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2253
2274 GCGATTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2333
2254 GCGATTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2313
2334 AAACCAAGTATGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2393
2314 CATTAAGACGACAGAGGATCTTGTGCTTCTGCGGATGATGATGATGATGATGATGATGATGAT 2273
2394 GGAATTCGACAAAGGAGAGCCGTCAGACGCGCGGATGATGATGATGATGATGATGATGATGATGAT 2453
2374 GGAATTCGACAAAGGAGAGCCGTCAGACGCGCGGATGATGATGATGATGATGATGATGATGATGAT 2433
2454 GCGATTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2513
2434 GCGATTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2493

2514 GCGAGTTTTCAGAGGCTCTGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2573
2494 GCGAGTTTTCAGAGGCTCTTAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2553
2574 GTCTTCAGAGGCTCAAGAGGCTCTGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2633
2554 GTCTTCAGAGGCTCAAGAGGCTCTGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2613
2634 ACCGCTCTGGAAGAAAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2693
2614 ACCGCTCTGGAAGAAAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2673
2694 GGCATGCGCAAGACAGGCGAGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2753
2674 GGAACCGGAAGGCGGCGCGAGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2733
2754 GACTCAGAGTCACTCCCGATTCACCACTCTGCGAGAAAGCTTCCAGAACCCCGGCTGCT 2813
2734 GACGCAAGTCACTCACTGACCCCGATTCGAGACGAGCAGCAGGCGCGGCTGCTGCT 2793
2814 GTGGGACTTCACTCAATGCTTCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2873
2794 CTGGGACTTCACTCAATGCTTCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2853
2874 GCGGACGAGTGGGATATGCTTCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2933
2854 GCGGACGAGTGGGATATGCTTCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2913
2934 AGAGTCACTCACTCAAGACCCCGATTCGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2993
2914 AGAGTCACTCACTCAAGACCCCGATTCGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2973
2994 AAGCAATCTTCAAGTCTTCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3053
2974 AAGCAATCTTCAAGTCTTCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3030
3054 ACCGCTCTGGAAGAAAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3113
3031 ACCGCTCTGGAAGAAAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3090
3114 CAGGACTCACTCAAGAAAGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3173
3091 CAAGACTCACTCAAGAAAGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3150
3174 AACATCAAGTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3233
3151 AACATCAAGTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3210
3234 ACCAGACGCTTCAAGTCTTCTGGAAGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3293
3211 ACCAGACGCTTCAAGTCTTCTGGAAGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3270
3294 GCGGACGAGGCTGCTCCCTCGGCTTCCCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3353
3271 GCGGACGAGGCTGCTCCCTCGGCTTCCCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3330
3354 TACCTGAGCTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3413
3331 TACCTGAGCTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3390
3414 TATTTCCCTTCTCAAGTCTTCTGGAAGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3473
3391 TATTTCCCTTCTCAAGTCTTCTGGAAGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3450
3474 GAAATGCTTTTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3533
3451 GAAATGCTTTTCAAGGAGTCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3510
3534 CTATGACCAATATGAT 3593
3511 CTATGACCAATATGAT 3570
3594 AACAGGACTTGTGTTTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3653

Db	3571	CAGTCAAGGCTTCAGTTTCTCTCAGGCCGAGACAGATGACATTTCGGAGCCACTCTAGAAC	3630
QY	3654	TGGCTACCTGGAACCTGTATTTCGACGACGCGCTTTTAAACAAAACAGACACAC	3713
Db	3631	TGGCTCTCTGGACCTCTGTTTACCCGACGACGAGATCAAAAGACATCTGCCGATPACAC	3690
QY	3714	AACAGCAATTTTAACTCGGACCTGGGCTCAAAATTTAACTCATTGGGCGTGAATCCATC	3773
Db	3691	AACGTGATATCTGCGGACTGGAGCTACAGTACCACTCTCAATTGGCAGACCTCTCG	3750
QY	3774	ATCAACCCCTGGCACTGCTATGAGCCCTCACACAAAGACGACGAAGACAAGTTCTTCCCATG	3833
Db	3751	GTTGAATCCGGGCCCGGCGCATGGCAAGCCACAGAGCAGATGAAGAAAGTTTTCCTCAG	3810
QY	3834	AGCGGTGTCATGATTTTGGAAAAGAGAGCCGCCGAGCTTCAAAACCTGCATTGGACAT	3893
Db	3811	AGCGGGGTTCATCTCTTGGGAGCAGAGGCTCAGAGAAAACAAATGTGGACATTGAAAG	3870
QY	3894	GTCATGATTAACAGAGGAAGAAATTTAAAGCAGCTAACCCGTGGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTAACAGAGGAAGAAATTCAGACACCAATCCCGTGGCTATCGAGACATAT	3930
QY	3954	GGGACCGGTGGAGTCAATTTCCAGAGCAGACAGACAGCCCTGGACCGGAGATGTGAT	4013
Db	3931	GGTTCTGTATCTACCAACCTCCAGAGGACAAACAGACAGGAGCTACCGCAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTAACCTGGCATGTGTGGCAAGTATAGACGTGTACTCGAGGCTCC	4073
Db	3991	ACACAGGCGCTTCTCCAGGAGATGCTCGGAGGACAGACAGATGTGTACTTTCAGGGGCC	4050
QY	4074	ATTGGGGCAAAATTCCTCACACAGATGGACACTTTCACCCGTCTCTTATGGGCGGC	4133
Db	4051	ATCTGGCAAAAGATTCCACACACGAGCGGACATTTTCACCCCTCTCTCATGGGTGGA	4110
QY	4134	TTTGGACTCAAGAAACCCGCGCTTCAGATCTCATCAAAAACAGCGCTGTTCCTGCAAT	4193
Db	4111	TTCGGACTTAAACACCCCTCTCCACAGATTCATCAAGAAACCCCGGATCTGGCAAT	4170
QY	4194	CCTCCGGGAGTTTTCAGCTACAAAGTTTGCTTATTCATCACCCCACTACTCCACAGGA	4253
Db	4171	CTTTCGACCACTTTCAGTGGGCAAAAGTTTCTCTTCATCAACACATATCTCCACGGGA	4230
QY	4254	CAAGTGAAGTGGAAATTTGAATGGAGCTGCGAGAAAGAAAACAGCAACGCTGGAATCC	4313
Db	4231	CAGGTCAAGCTGGAGATCGAGTGGAGCTGCGAAGAGAAAACAGCAACGCTGGAATCC	4290
QY	4314	GAAGTGACGATACACATCCAAATTAATGCAAAATCTGCCACGTTGATTTTACTGTGGACAC	4373
Db	4291	GAAATTCAGTACACTTCCCACTAACAAAGCTGTGTAATGTGGACTTATCTGTGGACACT	4350
QY	4374	AATGACCTTATATCTGAGCCTCGCCCCATTGGCACCCGTTACCTTACCCGCTCTGTA	4433
Db	4351	AATGGCGGTATTCAGAGCCTCGCCCCAATGGCACCAATACCTGCATCTGTAACTGTAA	4410
QY	4434	TTAGCGTTATTCATAAACCAGGTGATTCGTTTCAAGTTGAACCTTGTCTCTGTCCTT	4493
Db	4411	TTGCTGTGTTATTCATAAACCAGTTTAATTCGTTTCAAGTTGAACCTTGTCTCTGCGATTT	4470
QY	4494	CTTATCTTATCGGTTACCATGATTAATAGCTTACACATTAACCTGCTGTGGCGCTTGGCG	4553
Db	4471	TCCTTCTTATCTAGTTTTCATGGCTACGTAGATTAAGTATGACATGGCGGGTATATTAAC	4530
QY	4554	ATTAAGAAGCTTAACGTCATCGGGTTAACCTCTAGTATGAGATGGCCACTCCCTCTGGCG	4613
Db	4531	TACAAAGG-----ACCCCTAGTATGAGATGTGGCACTCTCTCTGCG	4574
QY	4614	CGCTTCGCTCGCTCGGTGGGCGCTGCGGACAAAGGTCGACAGCGGACGACCTCTGCTCT	4673
Db	4575	CGCTTCGCTCGCTCACTGAGGCGCGGCGACAAAGGTCGCCGACGCGCGGCTTTGGCCG	4634
QY	4674	GCCTGGCCCAACCGAGCTGACGAGCGCGCAGAGAGGGAAGTGGGCA	4718

Db	4635	GGCGGCTCAGTGAAGCGAGCGGCGGAGAGAGAGTGGCCAA	4679
RESULT 6			
ID	AAAF23750		
XX	AAAF23750 standard; DNA; 4679 BP.		
AC	AAAF23750;		
XX			
DT	28-MAR-2001 (first entry)		
XX			
DE	AAV2 DNA sequence.		
XX			
KM	AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;		
KW	atherosclerosis; sickle cell anaemia; thalassemia;		
KW	blood clotting disorder; diabetes; ss.		
XX			
OS	Adeno associated virus.		
XX			
PN	US6156303-A.		
XX			
PD	05-DEC-2000.		
PE	11-JUN-1997; 97US-0873168.		
XX			
PR	11-JUN-1997; 97US-0873168.		
XX			
PA	(UNIW) UNIV WASHINGTON.		
XX			
PI	Russell DW, Rutledge EA;		
XX			
DR	WPI; 2001-060164/07.		
XX			
PT	Adeno-associated virus serotype 6 and viral vector derived from it for		
PT	gene therapy of cystic fibrosis, cancer, acquired immunodeficiency		
PT	syndrome, sickle cell anemia, thalassemia and diabetes -		
XX			
PS	Claim 7; Fig 1; 50pp; English.		
XX			
CC	The present invention relates to adeno-associated virus serotypes. The		
CC	present sequence is the DNA sequence of one such serotype (AAV2). AAV2		
CC	can be used to construct AAV viral vectors for use in gene therapy for a		
CC	range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,		
CC	sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.		
CC	The AAV viral vectors have increased transduction efficiency of a		
CC	particular host cell as the AAV virion containing the AAV vector genome		
CC	can be modified to express a capsid protein of an AAV serotype that		
CC	transduces the selected host cell.		
XX			
SQ	Sequence 4679 BP; 1197 A; 1263 C; 1255 G; 964 T; 0 other;		
Query Match	65.8%; Score 3106.2; DB 22; Length 4679;		
Best Local Similarity	80.2%; Pred. No. 0;		
Matches 3789; Conservative	0; Mismatches 883; Indels 53; Gaps 10.		
QY	1 TTGGCCAATCCCTCTCTGCGGCTCGCTCGCTCGGTGGGCGCTCGGACCAAGGTCCG	60	
DB	1 TTGGCCAATCCCTCTCTGCGGCTCGCTCGCTCGGTGGGCGCTCGGACCAAGGTCCG	60	
QY	61 AGAGCGCAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGCGGAGAGGAGTG	120	
DB	61 CGAGCGCGGCGCTTGGCCCGGCGGCTCGATGAGCGAGCGCGCGAGAGGAGTG	120	
QY	121 GCGAATCCATCAGTGAAGGAGTATCGGAGAGCGCTCCACAGCTGCGCGCTGAGCGCTGA	180	
DB	121 GCGAATCCATCAGTGAAGGAGTATCGGAGAGCGCTCCACAGCTGCGCGCTGAGCGCTGA	180	
QY	181 CGTAAATTAACATCAGTGAAGG--GAGTGTCTCTATTAAGTGTACGTGAGTCTTTTG	237	
DB	164 CGTAAATTAACATCAGTGAAGGAGTCTCTATTAAGAGTGTACGTGAGTGT--TTTTC	222	
QY	238 GACATTTTGGCAGACCAAGTGGCCATTTAGGATATATATGCGCGAGTGAAGCAGAGAT	297	

Db 223 GACATTTTGGACACACATGTGTCAACGCTGGGTATTTAAAGCCGAGTGAACGACGAGGT 282
Qy 298 CTCATTTTGG-ACGCGAATTTGAGAGACGACGACCCATGCGGGCTTCTAGACATCG 356
Db 283 CTCATTTTGGAGGGAGGTTTGAACGCGACGCGCATGCGGGGTTTACAGATG 342
Qy 357 TGATCAAGGTCGAGGACCTGAGACGACCTGCGGGGATTTGACTGCTGTTTGA 416
Db 343 TGATTAAGTCCCGACGACCTTGAACGGGACATCTGCCGACATTTCTGACGCTTTGTA 402
Qy 417 GCTGGTGGCCGAGAGGATGGAGCTGCCCCCGGATTTCTGACATGATCTGAATCTGA 476
Db 403 ACTGGGTGGCCGAGAGGATGGAGTGGCCGACGATTTCTGACATGATCTGAATCTGA 462
Qy 477 TTGAGCAGGCAACCTTGAACCTGTGGCCGAGAGCTGACGCGGACTTCTGTGCTCAATGCG 536
Db 463 TTGAGCAGGCAACCTTGAACCTGTGGCCGAGAGCTGACGCGGACTTCTGTGAGGAATGCG 522
Qy 537 GCCGCGTGAAGTGAAGCCCGGAGGCGCTCTTCTTGTTCAGTTCGAGAGGCGAGTCTCT 596
Db 523 GCCGTGTGAAGTGAAGCCCGGAGGCGCTCTTCTTGTTCAGTTCGAGAGGCGAGTCTCT 582
Qy 597 ACTTCAACCTTCATATTTCTGTGAGAGACACGCGGGGTCAATTCATGTGCTGGGCGCT 656
Db 583 ACTTCAACATGCACTGTCTGTGGAACCAACGGGGGTGAATTCATGTGTTTGGGACGTT 642
Qy 657 TCTGATGTCAGATTAGGAGCAAGCTGTGACGACATCTACCGGGGATTCGAGCGCAACC 716
Db 643 TCTGAGTCAAGATTGCGAATAACTGATTGAGAGATTTACCGGGGATTCGAGCGCACTT 702
Qy 717 TGCCCACTGGTGGCGGTGACCAAGACGCGTAAATGGCGCGCGAGGGGAGCAAGGTGG 776
Db 703 TGCCCAATGTTGTCGGGTCAAAAGACCAAGAAATGGCGCGGAGCGGAGCAAGGTGG 762
Qy 777 TGAAGAGTGTACATCCCACTACTCTCTGACCAAGACTGACCGGAGCTGCAATGGG 836
Db 763 TGAATGAGTGTACATCCCACTACTCTCTGACCAAGACTGACCGGAGCTGCAATGGG 822
Qy 837 CGTGAATCAATGAGAGAGTATTAAGCGCTGTTTGAACCTTGCGGAGGGCAAGCGGC 896
Db 823 CGTGAATCAATGAGAGAGTATTAAGCGCTGTTTGAACCTTGCGGAGGGTAAACGCT 882
Qy 897 TCGTGGCGGACGACCTGACCCGACGTCAGCCGAGAGCAGAGCAAGAGAGATCTGA 956
Db 883 TCGTGGCGGACGACCTGACCCGACGTCAGCCGAGAGCAGAGCAAGAGAGATCTGA 942
Qy 957 ACCCCAATTTGACGCGCTGTCAATCGGTCAAAAACCTCGCGGCTCAATGAGAGCTGG 1016
Db 943 ATCCCAATTTGATGCGCGGTGATCAGATCAAAAACCTTCAGCAGGTAATGAGAGCTGG 1002
Qy 1017 TCGGGTGGCTGGTGAACCGGGGACATCACTCGAGAGAGAGATCCAGAGAGACAGG 1076
Db 1003 TCGGGTGGCTGGTGAACAGGGGATTAATCTCGAGAGAGAGATCCAGAGAGACAGG 1062
Qy 1077 CTTGTGATCTCTTCAACGCGCTTCAACCTGCGGTGTCAGATCAAGGCGCTGCG 1136
Db 1063 CTTGTGATCTCTTCAACGCGGTGTCAGATCAAGGCGCTGCGGTGCGCTTGG 1122
Qy 1137 ACAATGCGGCAAGATCAAGCGCTGACCAATCGCGCGCTCACTCTGTGAGCGCCG 1196
Db 1123 ACAATGCGGCAAGATTAATGAGCTGATTAACCGCGCGCTCACTCTGTGAGCGCG 1182
Qy 1197 CTGCGCGCGGAGATTAACCAACCGCATCTACCGCATCTGAGAGTGAAGCGGTACG 1256
Db 1183 AGCCGCGGAGAGATTTCCAGCAATCGGATTTAATAATTTGAACTTAACCGGTACG 1242
Qy 1257 AACCTGCTAGCGCGCTCGCTCTTCTCGGCTGGGCGCAGAAAAAGTTTCGAGAGCGCA 1316
Db 1243 ATCCCAATATGCGGCTTCGCTCTTCTGAGATGGGCGCAGAAAAAGTTTCGAGAGAGGA 1302
Qy 1317 ACACCATGTGGCTGTTGGGCGGCGCACGAGCGGAGAGACGAATCGCGGAGCGCATCG 1376
Db 1303 ACACCATGTGGCTGTTGGGCGGCGCATGACCGGAGAGACGAATCGCGGAGCGCATGAG 1362

Qy 1377 CCGACGCGGCGCTTCTACGCGTGGTCAACGTGACCAATGAGAACTTTCCCTGAATG 1436
Db 1363 CCGACATGCGCTTCTACGCGGTGTAACGTGAACTGAACTGAACTTTCCCTTCAAG 1422
Qy 1437 ATTGCGTCAAGATGTGTATCTGTGTGAGAGAGGAGCAATGACGCGCAAGTCTGTG 1496
Db 1423 ACTGTGTCAAGATGTGTATCTGTGTGAGAGAGGAGAGATGACCGCAAGTCTGTG 1482
Qy 1497 AGTCCGCAAGGCGCATTTCTGCGCGGACGAGGTGCGCTGTGACCAAAATGCAATGCT 1556
Db 1483 AGTCCGCAAGGCGCATTTCTGAGAGAGAGCAAGGTGCGCTGTGACCAAAATGCAAGCT 1542
Qy 1557 CCGCGCAGATGACCCCAACCCCGGTATGCTCACTTCAACACCAACATGTGCGCGTGA 1616
Db 1543 CCGCGCAGATGACCCCAACCCCGGTATGCTCACTTCAACACCAACATGTGCGCGTGA 1602
Qy 1617 TTGACGGGACAGACCACTTGTGAGACCAAGAGCGGTTGACAGACCGGATGTTCAAT 1676
Db 1603 TTGACGGGACCTCAAGACCTTGTGAACACAGCAGCGGTTGCAAGACCGGATGTTCAAT 1662
Qy 1677 TTGAACTCACCCGCGCTGTGAGATGACTTTGGCAAGGTGACAAAGCAGAAAGTCAAG 1736
Db 1663 TTGAACTCACCCGCGCTGTGATCACTTTGGGAAAGGTCAACAGCAGAAAGTCAAG 1722
Qy 1737 AGTTCTTCCGCTGGGCGCAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAGAA 1796
Db 1723 ACTTTTCCGCTGGGCGCAAGAGATCACTGCTGTGAGGTGAGCATGATTTCTAGTCAAA 1782
Qy 1797 AGGGTGAAGCCAAAGAGACCCGCGCGAGTACGCGGATTAAGCGAGCCCAAGCGGG 1856
Db 1783 AGGGTGAAGCCAAAGAGACCCGCGCGAGTACGCGGATTAAGCGAGCCCAAGCGGG 1842
Qy 1857 CTTGCCCTCAAGTCGCGGATTCATCGACGTCAGACCGGAGAGAGTCTCCGCTGGAATTTG 1916
Db 1843 TGGCGAGTGAATGGCGACGACATCAAGTCAGACGCGGA--AGCTTCATCAACTACG 1899
Qy 1917 CCGACAGGTACCAAAACAAATGTTCTCGTACGCGGGCATGCTTACAGATCTGTTCCCT 1976
Db 1900 CAGACAGGTACCAAAACAAATGTTCTCGTACGCGGGCATGCTTACAGATCTGTTCCCT 1959
Qy 1977 GCAAGACATGAGAGATGAATCAGAAATTTCAACTTTGCTTACGACGCGAGAG 2036
Db 1960 GCAAGACATGAGAGATGAATCAGAAATTTCAACTTTGCTTACGACGAGAGAG 2019
Qy 2037 ACTGTTCAAGTCTTCCCGCGGTGACAGATCTCAACCGGT---GTCAGAAAGGA 2093
Db 2020 ACTGTTCAAGTCTTCCCGCGGTGACAGATCTCAACCGGT---GTCAGAAAGGA 2076
Qy 2094 CGTATCGGAAACTCTGTGCAATTCATGCTGCTGGGCGGGCTCCGAGATTTGCTTGT 2153
Db 2077 CGTATCGGAAACTGTGCTATCATTCATTCATGAGGAAGGTGCCAG---GCTTGCA 2133
Qy 2154 CCGCTGCGATCTGTGCAACGTGAGCTGAGTATCTGTGTTCTGAGCAATTAATGACTT 2213
Db 2134 CTTGCTGCGATCTGTGCAATTCATGCTGATGATGATGATCTTGTGAACAAATTAATGATTT 2193
Qy 2214 AAACAGGTATGTGCTGCGGATGTTATCTTCAAGTTGGTGTGAGAGCAACTCTGAG 2273
Db 2194 AAATCAGGTATGTGCTGCGGATGTTATCTTCAAGTTGGTGTGAGAGCAACTCTGAG 2253
Qy 2274 GGCATTCGAGTGTGAGGACTTGAACCTGAGCCCGAGAGCCCAAGCCCAAGCCCAAG 2333
Db 2254 GGAATTAAGACATGTGTGAGAGCTCAAACTGCGCCACCAACCAACCAAGCCCGAGAGCGG 2313
Qy 2334 AAGCAGAGCAGCGCGGAGTCTGTGCTTCTGTGCTTCAAGTACTCGACCTTCAAC 2393
Db 2314 CATTAAGACAGCAGGAGGCTTGTGCTTCTGTGCTTCAAGTACTCGACCTTCAAC 2373
Qy 2394 GGAATTCGAGGAGGAGCCGCTCAACGCGCGGAGAGCGACCGCGCTCGAGCAGAGAG 2453
Db 2374 GGAATTCGAGGAGGAGCCGCTCAACGAGGAGAGCGCGCGCTCGAGCAGAGAG 2433

QY 2454 GCCTACAGCAGCAGCTCAAGCCGGGTGACATCCGTACCTGCGGTAAACACAGCCGAC 2513
Db 2434 GCCTACAGCAGCAGCTCAAGCCGGGTGACATCCGTACCTGCGGTAAACACAGCCGAC 2493
QY 2514 GCGGAGTTTCAAGAGCGCTCTGCAAGAAAGATACGTTTGGGGGCAACCTCGGCGAGCA 2573
Db 2494 GCGGAGTTTCAAGAGCGCTCTGCAAGAAAGATACGTTTGGGGGCAACCTCGGCGAGCA 2553
QY 2574 GTCTTCAGGCGAAGAGCGGGTCTCGAACCTCGGTGCTGCTGAGAGAGGGCTAAG 2633
Db 2554 GTCTTCAGGCGAAGAGAGGGTCTGAGAACCTCGGTGCTGCTGAGAGAGGGCTAAG 2613
QY 2634 ACCGCTCTGAGAAAGAAACGTCCGCTGAGAGAGTCCGACAGAGCCGAGACTCCTCTCG 2693
Db 2614 ACCGCTCTGAGAAAGAAAGGGCTGAGAGACTCCTGCTGAGAGCCGAGACTCCTCTCG 2673
QY 2694 GGCATCGGCAAGACAGGCGCAGAGCCCGCTAAGAAAGAGACTCAATTTTGGTCAAGTGGC 2753
Db 2674 GGAACCGGAAAGGCGGCGCAGAGCTGCAAGAAAGAAATTTGATTTGGTCAAGTGGG 2733
QY 2754 GACTCAGAGTCAAGTCCCGGATCCACAACTCTCGGAGAACCTCGAGAACCCCGCTGCT 2813
Db 2734 GAGCGACACTCAAGTCAAGTCCCGGATCCCGGAGAACCTCGAGAACCCCGCTGCTGCT 2793
QY 2814 GTGGGACCTTACTCAATGCGCTTCAGGCGGTGCGCACCAATGGCAGACATTAAGAGGC 2873
Db 2794 CTGGGACCTTAAATGAGTGGCTTACAGGAGTGGCGCACCAATGGCAGACATTAAGAGGC 2853
QY 2874 GCCGACGAGTGGGTAAATGCTTCAGAAATTTGGCATTTCCATGAGTGGGCGAC 2933
Db 2854 GCCGACGAGTGGGTAAATTTCTCCCGGAAATTTGGCATTTCCATGAGTGGGCGAC 2913
QY 2934 AGAGTCACTACACACAGACCCGACCTGGGCTTGGCCCACTACATTAACACTTAC 2993
Db 2914 AGAGTCACTACACACAGACCCGACCTGGGCTTGGCCCACTACATTAACACTTAC 2973
QY 2994 AAGCAATCTCAGTGGCTTCAAGCGGGGCGAGCAACAGCAACATCTCTCGGCTACAGC 3053
Db 2974 AAGCAATTTCCAGCCATCA---GAGCTCCAGAACCAATCACTTCTTGGCTACAGC 3030
QY 3054 ACCCCCTGGGGGTATTTTGAATTTCAACAGATTCACATGCGCACTTTTCACACGTGACTGG 3113
Db 3031 ACCCTTGGGGGTATTTTGAATTTCAACAGATTCACATGCGCACTTTTCACACGTGACTGG 3090
QY 3114 CAGGACTCATCAACAACTATGGGGATTCGCGCCCAAGAGACTCAACTTCAACTTTC 3173
Db 3091 CAAAGACTCATCAACAACTATGGGGATTCGCGCCCAAGAGACTCAACTTCAAGCTCTTT 3150
QY 3174 AACATCAAGTCAAGAGGTCAAGACGATGATGCGTCAACACATCGCTAATAACTT 3233
Db 3151 AACATTCATTCAAAGAGGTCAAGACGATGATGCGTCAACACATTCGCTAATAACTT 3210
QY 3234 ACCAGACGAGTTCAAGTCTTCTCGGACTCGGAGTACCAAGTTCGATCGTCT 3293
Db 3211 ACCAGACGAGTTCAAGGTTTACTGACTCGGAGTACCAAGTTCGATCGTCTCGGCTCG 3270
QY 3294 GGGCACAAGGAGTCTCTCTCGGCTTCCGGCGGAGGCTTCAATGATTCGCGAATACGCG 3353
Db 3271 GGGCACAAGGAGTCTCTCTCGGCTTCCGGCGGAGGCTTCAATGATTCGCGAATACGCG 3330
QY 3354 TACCTGACGCTCAACATGGCAGCAAGCGTGGGAGCTTCACTCTTTTACCTGAGAA 3413
Db 3331 TACCTGACGCTCAACATGGGAGTCAAGGAGTCAAGGAGCTTCTTCACTTTTACCTGAG 3390
QY 3414 TATTTCCCTTCTAGATGCTGAGAACGGGCAACACTTTTACCTTCAAGCTACACTTGG 3473
Db 3391 TACTTTCCTTCTAGATGCTGCGTAAACGGAAACACTTTTACCTTCAAGCTACACTTGG 3450
QY 3474 GAGTGGCTTTCACAGCAGCTACGCGCAACAGCAGAGCTGAGACCGGCTGATGATTCCT 3533
Db 3451 GAGTGGCTTTCACAGCAGCTACGCTACAGCAGAGCTGAGACCGTCTCAAGATTCCT 3510
QY 3534 CTGATGACCAATACCTGTATTAACCTGAAACAGAACTCAATATCATGTCGGAGTGGCAA 3593

Db 3511 CTGATGACCAATACCTGTATTAACCTGAGAGAAACAACTCAAGTGAACACCAACG 3570
QY 3594 AACAGGACTTGTCTGTTTACCGTGGCTTCCAGCTGGCATGTCTTCAAGCCCAAAAC 3653
Db 3571 CAGTCAAGGCTTCAAGTTTCTCAGGCGGAGCCAGTGAACATTCGGGACCACTGCTAGAAC 3630
QY 3654 TGGCTACCTGGACCCGTTTATGGGAGAGCGGCTTCTTAAACAAACACAGCAACAC 3713
Db 3631 TGGCTACCTGGACCCGTTTATGGGAGAGCGGCTTCTTAAACAAACACAGCAACAC 3690
QY 3714 AACAGCAATTTTACTCTGAGCTGTGCTTCAAAATATTAACCTCAATGAGGCTGATCATC 3773
Db 3691 AACAGCAATTTACTCTGAGCTGTGAGCTGAGCTCAACATTAACCTCAATGAGGCTGAT 3750
QY 3774 ATCAACCTGGGACTGCTGTATGCTTCAACAAAGACGAGAGCAAGTCTTCCATG 3833
Db 3751 GTGAATCCGGGCGCGGCAATGGCAAGCCCAAGAGAGATGAAGAAAGTTTCTTCCAG 3810
QY 3834 AGCGGTGCAATGTTTTGAGAAAGAGAGCGCGGCTTCAACATGATGAGCAAT 3893
Db 3811 AGCGGAGTCTCATCTTTGAGAAAGAGGCTCAAGAGAAACAAATGTGACATTTGAAAG 3870
QY 3894 GTCATGATTTACAGACGAGAGAAATTTAAAGCCATTAACCTGTGGCCACCGAAAGATT 3953
Db 3871 GTCATGATTTACAGACGAGAGAAATTTACAGAGCAACCAATCCCTGGCTACGAGCAAT 3930
QY 3954 GGGACCTGGCACTCAATTTCCAGAGCAGACAGACCTTGGCAACCGGAGATGTCAT 4013
Db 3931 GGTTCGTATCTTCAACCTCCAGAGAGGCAACAGCAAGCTACCCGAGATGTCATC 3990
QY 4014 GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATGAGAGCTGATCTGCGAGGTCCT 4073
Db 3991 ACACAGGAGTCTTCAAGCATGTGTGGCAAGAGATGATGATCTTCAAGGAGGCTCC 4050
QY 4074 ATTTGGGCAAAATTTCTTCAACAGATGAGCACTTCAACCCGCTCTCTTATGAGGCGC 4133
Db 4051 ATCTGGGCAAAAGTTTCAACAGAGGAGCAATTTTCAACCCCTCTCTCTATGAGTGA 4110
QY 4134 TTTGACTCAAGAACCGGCTCTCTAGATCTTCATCAAAACACAGCTTCTGCGAAT 4193
Db 4111 TTGCGACTTAAACACCTCTCTTCAACAGATTTCAATCAAGAACACCCGGTACCTGGAAT 4170
QY 4194 CTTCCGGCGGAGTTTCACTTCAAGTTTGTCTTCAATCAACCAATATCTTCAAGGA 4253
Db 4171 CTTTCAACACCTTCAAGTCTGCGAAGTTTGTCTTCAATCAACAGATATCTTCAAGGA 4230
QY 4254 CAAAGTGTGGAATTTGAATGGGAGCTGCAAGAAAGAAACAGCAAGGCTGGAATCC 4313
Db 4231 CAGTCAAGCTGAGATGAGTGGAGCTGCAAGAGAAACAGCAAGGCTGGAATCC 4290
QY 4314 GAAGTCACTACATCAATTAATGCAAAATCTGCCAGTTGATTTTACTGTGACAA 4373
Db 4291 GAATTCAGTACACTTCAACTTAAACAGTCTGTTAATGTGACTTACTGTGACACT 4350
QY 4374 AATGACTTATTAAGAGCTTGGCCCATTTGGCAACCCGTACTTACCTTCCCTGTAA 4433
Db 4351 AATGAGTATTAAGAGCTTGGCCCATTTGGCAACCAATGACTGACTGTATATGTAA 4410
QY 4434 TTAAGTATTAATCAATTAACCGGTTGATGTTTCAAGTGAACCTTGTGCTCGTCC 4493
Db 4411 TTGCTTGTATTAATCAATTAACCGTTTATTTGTTTCAAGTGAACCTTGTGCTCGTAT 4470
QY 4494 CTATATTAATCGGTTAATCAATGTTTATGCTTACATTAATGCTTGTGCTTGGC 4553
Db 4471 TCTTTCTTATCTAGTTTCAATGCTTACGATGATGATGATGATGATGATGATGAT 4530
QY 4554 ATTAAGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4613
Db 4531 TACAAAGG-----ACCCCTGATGATGATGATGATGATGATGATGATGATGAT 4574
QY 4614 CGCTCGCTCGCTCGGTTGGGCGCTGCGGACCAAGGTCGCAAGCGGACAGCTGCTT 4673

Db 4575 CGCTGCTGCTCACTGAGCGCGGCGACCAAGGTCGCCCGCCGCGGCTTTGGCCG 4634
Qy 4674 GCGGCCCCCAGCGAGCGAGCGCGAGAGGAGTGG3CAA 4718
Db 4635 GCGCGCTCTGAGTGAAGCGAGCGCGAGAGAGGAGTGG3CAA 4679

RESULT 7

AA09008
ID AA09008 standard; DNA; 4680 BP.

AA09008;
AC AA09008;

DT 05-JUL-1996 (first entry)
XX

DE Wild-type adeno-associated virus 2 genome.

XX adeno-associated virus 2; AAV2; inverted terminal repeat; viral vector;
KW DNA delivery; immunodeficiency virus protein; immunity; human; simian;
KM neurological disorder; therapy; ss.

XX Adeno-associated virus.

XX MO9534670-A2.

XX 21-DEC-1995.

XX 06-JUN-1995; 95WO-US07178.

XX 06-JUN-1994; 94US-0254358.

XX (CHIL-) CHILDRENS HOSPITAL INC.

XX Johnson PR;

XX WPI; 1996-049697/05.

XX Recombinant adeno-associated virus genome contg. protein encoding
PT DNA - flanked by inverted terminal repeats, for use in vaccines or
PT for treatment of neuro-degenerative disease

XX Example 1; Page 27-29; 44p; English.

XX The present sequence is that of the wild-type adeno-associated virus
CC (AAV) 2, a replication-deficient parvovirus, cis-acting sequences
CC directing viral DNA replication (ori), encapsidation/packaging (pRI) and
CC host cell chromosome integration (int) are contained within the ITRs
CC (inverted terminal repeats of 145 nucleotides). When AAV infects a
CC human cell, the viral genome integrates into chromosome 19 resulting in
CC latent infection of the cell. Prodn. of infectious virus does not occur
CC unless the cell is infected with a helper virus (e.g., adenovirus or
CC herpes- virus). AAV possesses unique features that make it attractive as
CC a vector for delivering foreign DNA to cells. A vector including a
CC recombinant AAV genome contg. a simian immunodeficiency virus (SIV) rev
CC and envelop (gpl60) gene cassette was constructed from an existing plasmid
CC designated pnb201. The plasmid contains a modified wild-type AAV2
CC genome. Specifically, two XbaI sites were added via linker addition at
CC sequence positions 190 and 4484. These sites are internal to 191 bp ITRs
CC which include the 145 bp ITRs of the AAV genome. The insertion of these
CC sites allows the complete removal of the internal 4.3 kb fragment contg.
CC the AAV rep-cap genes upon XbaI digestion of the plasmid. (See also
CC AA09009-10).

XX Sequence 4680 BP; 1198 A; 1265 C; 1254 G; 963 T; 0 other;

XX Query Match 65.3%; Score 3082.4; DB 17; Length 4680;

XX Best Local Similarity 80.0%; Pred. No. 0;

XX Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

Qy 1 TTGGCCACTCCCTCTGCGCGCTGCGCTGCGTGGGCGCTGCGCAAAAGTCCGC 60
Db 1 TTGGCCACTCCCTCTGCGCGCTGCGCTGCGTGGGCGCTGCGCAAAAGTCCGC 60

Qy 1 TTGGCCACTCCCTCTGCGCGCTGCGCTGCGTGGGCGCTGCGCAAAAGTCCGC 60
Db 1 TTGGCCACTCCCTCTGCGCGCTGCGCTGCGTGGGCGCTGCGCAAAAGTCCGC 60

Qy 61 AGACGGCAGAGCTCTGCTCTGCGCGCCCAACCGAGCGAGCGAGCGCAGAGAGAGTG 120
Db 61 CGAGCGCGCGCGCTTTGCGCGCGCGCGCTCTAGTAGGAGGAGCGCGCAGAGAGAGAGTG 120
Qy 121 GCGAATCTCCATCTGAGGGGTAAATCGGAAAGCCCTCCACGCTGCCGCTGACGCTGA 180
Db 121 GCCAATCTCCATCTGAGGGGT-----TCTGAGAGGGGTGAGTGTGA 163
Qy 181 CGTAATTAAGTATGAGG---GAGTGTCTGTATTAAGTGTGACGTAGTCTTTTG 237
Db 164 CGTAATTAAGTATGAGGTTAGGAGGTCCTGTATTAAGTGTGACGTAGTGTG-TTTTG 222
Qy 238 GACATTTTGGACACCAAGCGGCTTAAAGGTATATGAGCGAGTGAAGCAGAGAT 297
Db 223 GACATTTTGGACACCAAGTGTGTGACGTGGGTATTTAAGCCCGAGTGAAGCAGAGG 282
Qy 298 CTCGATTTTG-ACCGGAAATTTGAACGAGCAGACCATGCCGCTTCTAGAGATG 356
Db 283 CTCGATTTTGAACCGGAGGTTTGAACGCGACGCCCATGCGGGGTTTACGAGATTG 342
Qy 357 TGATCAAGGTGCGAGCGACCTGAGCAGACCTGCGGGCATTTCTGACTGTTTGTGA 416
Db 343 TGATTAAGTCTCCAGGACCTTGACGGCATCTGCCCGCATTTCTGACAGCTTTGTGA 402
Qy 417 GCTGGGTGCGCGAAGAGATGGAAGTGTGCGCCCGGATTTCTGACATGATCTGAATCTGA 476
Db 403 ACTGGGTGCGCGAAGAGATGGAAGTGTGCGCCCGCATTTCTGACATGATCTGAATCTGA 462
Qy 477 TTGAGCAGGACCCCTGACCGTGGCCGAGAGAGTGCAGCGCCGCTTCTGTGCTCAATGCG 536
Db 463 TTGAGCAGGACCCCTGACCGTGGCCGAGAGAGTGCAGCGCCGCTTCTGACGGAATGCG 522
Qy 537 GCGCGGTGAGTAAAGCCCGAGGCGCTCTTTGTGTGAGTTCGAGAGGCGAGTCTCT 596
Db 523 GCGGTGAGTAAAGCCCGAGGCGCTCTTTGTGTGAGTTCGAGAGGCGAGTCTCT 582
Qy 597 ACTTCAACTCTCATTTTGTGTGAGAGCAACGCGGGGTCAATTCATGTGTGGGCGCT 656
Db 583 ACTTCAACTCTCATTTTGTGTGAGAGCAACGCGGGGTCAATTCATGTGTGGGCGCT 642
Qy 657 TCTGAGTCAAGTTAGGAGACAGCTGTGCAACATCTACCGGGGATTCGAGCCGACCC 716
Db 643 TCTGAGTCAAGTTAGGAGAGCACTGATTCAGGAATTTACCGGGGATTCGAGCCGACCTT 702
Qy 717 TGCCCAACTGTTGCGGAGTACCAAGACGCTATGCGCGCGGAGGGGGAACAAGTGTG 776
Db 703 TGCCCAACTGTTGCGGAGTACCAAGACGCTATGCGCGCGGAGGGGGAACAAGTGTG 762
Qy 777 TGAGCAGTGTCTATCTCCCACTACTCTGCGCAAGACTAGCGGAGCTGAGTGG 836
Db 763 TGAGTGTGTCTATCTCCCACTACTCTGCGCAAGACTAGCGGAGCTGAGTGG 822
Qy 837 CGTGAATCAATGAGAGAGTATATAGGCGCTTTGAACCTGCGCGAGCGCAACCGC 896
Db 823 CGTGAATCAATGAGAGAGTATATAGGCGCTTTGAACCTGCGCGAGCGCAACCGC 882
Qy 897 TCGTGCGCAGCAGCTGACCCAGCTGAGCAGAGCCAGAGCAGAGCAAGAGAGATCTGA 956
Db 883 TCGTGCGCAGCAGCTGACCCAGCTGAGCAGAGCCAGAGCAGAGCAAGAGAGATCTGA 942
Qy 957 ACCCCATTTGAGCGCGCTGTATCCGCTCAAAAACCTTCGCGCGCTACATGAGCTGG 1016
Db 943 ATCCCATTTGAGCGCGCTGTATCCGCTCAAAAACCTTCAGCAGGATCATGAGACTGG 1002
Qy 1017 TCGGAGTGTGTGAGACCGGGGATCACTCCGAGAGAGAGTGTATCCAGAGAGACAG 1076
Db 1003 TCGGAGTGTGTGAGACCGGGGATTAATCTCGAGAGAGAGTGTATCCAGAGAGACAG 1062
Qy 1077 CCGTGAATCAATCTTCAACGCGCTTCAACCTGCGGCTCCGAGATCAAGGCGCTGTGG 1136
Db 1063 CCGTGAATCAATCTTCAACGCGCTTCAACCTGCGGCTCCGAGATCAAGGCGCTGTGG 1122
Qy 1137 ACAATGCCGCAAGATCATGCGCTGACCAAAATCCGCGCCGACTGACTGTGAGGCCCCG 1196

Db 1123 ACATGCGGGAAAGATTATGAGCTGCTAAACCGGCCCCGACTACCTGTGGGCGAG 1182
 QY 1197 CTCCGCCCGGGGACATTAAAAACGAGCTCTACCGCATCTCGAGCTGAAGGGCTAG 1256
 Db 1183 AGCCCGGTGGAGGACATTTCCAGCAATCGATTATTAATTTTGGAACTTAAAGGGGTAG 1242
 QY 1257 AACCTGCTACGCGCGCTCCGCTCTTTCTCGGCTGGGCGGAGAAAGGTTGGAGAGCG 1316
 Db 1243 ATCCCAATATGGGGCTTCCGTCTTTCTGGGATGGGCGAGAAAAGTTGGCAAGAGA 1302
 QY 1317 ACAACATCTGGCTGTTTGGGCGGCAACGCGGCAAGCAACATCGCGAAGCAATCG 1376
 Db 1303 ACACATCTGGCTGTTTGGGCTGCAACTCGGGAAGACCAACATCGCGAGGACATAG 1362
 QY 1377 CCACGCGGCTGCTCTCTACGCTGCTGCACTGAGCAATGGAATCTTCCCTTCAATG 1436
 Db 1363 CCACACCTGGCTCTCTACGCTGCTGCACTGAGCAATGGAATCTTCCCTTCAACG 1422
 QY 1437 ATGCGCTGCAAGATGATGATCTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1496
 Db 1423 ACTGTGTCAGCAAGATGATGATCTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 1482
 QY 1497 AGTCCGCAAGGCACTTCTGCGCGAGCAAGGTCGCGCTGCAACAAAGTGCAGTCTG 1556
 Db 1483 AGTCCGCAAGGCACTTCTGCGAGGAAGCAAGGTCGCGCTGCAACAAAGTGCAGTCTG 1542
 QY 1557 CGGCGGAGTCCGACCCGCGCTGATGCTCACTTCCAAACCAACATGTGGCGCTGA 1616
 Db 1543 CGGCGGAGTCCGACCCGCGCTGATGCTCACTTCCAAACCAACATGTGGCGCTGA 1602
 QY 1617 TTGACGGGAACAGACCACTTGGAGACAGGAGCGGTCGAGGACCGGATGTTCAAT 1676
 Db 1603 TTGACGGGAACAGACCACTTGGAGACAGGAGCGGTCGAGGAGCGGATGTTCAAT 1662
 QY 1677 TTGAACTCAACCGCGCTGAGACATGACTTTGGCAAGGTGACAAAGCAAGAGTCAAG 1736
 Db 1663 TTGAACTCAACCGCGCTGAGACATGACTTTGGCAAGGTGACAAAGCAAGAGTCAAG 1722
 QY 1737 AGTCTTCCGCTGGGCGGAGATCACTGACCGAGGTCGCTGAGTCTTACGTCAAG 1796
 Db 1723 ACTTTTCCGCTGGGCGGAGATCACTGAGGTGAGGAGCATGAAATCTTACGTCAAG 1782
 QY 1797 AGGCTGAGGCAACAAAGACCGCGCGGATGACGCGATTAAGCGAGCCCAAGCGGG 1856
 Db 1783 AGGCTGAGGCAACAAAGACCGCGCGGATGACGCGATTAAGGTAGCCCAAGCGGG 1842
 QY 1857 CTTGCCCTCACTGCGGATCACTGACGTCAGACGCGGAAGAGCTCCGCTGAGCTTTG 1916
 Db 1843 TGGCGAGTCACTTGGCGAGCATCACTGACGCGGGA---AGCTTGATCACTAG 1899
 QY 1917 CGGACAGGTACCAAAACAAATGTTCTGCTACGCGGCGATGCTTCAAGTCTGTTCCCT 1976
 Db 1900 CAGACAGGTACCAAAACAAATGTTCTGCTACGCGGCGATGCTTCAAGTCTGTTCCCT 1959
 QY 1977 GCAAGCATGCGAGGAATGATCAGAAATTTCAACATTTGCTTCAAGCAAGGAGGAG 2036
 Db 1960 GCAAGCATGCGAGGAATGATCAGAAATTTCAACATTTGCTTCAAGCAAGGAGGAG 2019
 QY 2037 ACTGTTCAAGTCTTCCCGGCGGTGACAAATCTCAACCGGT---CGTCAAGGAAGGA 2093
 Db 2020 ACTGTTCAAGTCTTCC---CGTCAAGATCTCAACCGGTGCTGTCGCAAAAAG 2076
 QY 2094 CGTATCGAAGACCTGTGCGATCACTGCTGCGGCGGAGTCCCGAGATGCTTCTG 2153
 Db 2077 CGTATCGAAGACCTGTGCTATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2133
 QY 2154 CGGCTGCGATCTGCTCAACGCTGGAACCTGATGATGTTTCTGACCAATTAATGACT 2213
 Db 2134 CTGCTGCGATCTGCTCAACGCTGGAATTTGATGATGCTGCTTCAACATTAATGATTT 2193
 QY 2214 AACCAAGTATGCTGCGATGATTTCTTCAAGATTTGCTGAGGACCACTCTCTGAG 2273

Db 2194 AATCAGTATGCTGCCAGTGTATCTTCCAGATTTGCTGAGGACCACTCTCTGA 2253
 QY 2274 GGCAATTCGAGTGTGGGAGCTTGAACCTGAGGCGGCAAGCCCAAGGCAACAGCA 2333
 Db 2254 GGAATTAAGCAGTGTGGAGGCTCAACCTGCGCCACCAACCAAGCCGAGGCGG 2313
 QY 2334 AAGCAGAGCAGCGCGGAGTGTGCTTCTGCTGCAATGATCTCGAATCTTCAAC 2393
 Db 2314 CATTAAGAGCAGCAGGAGTGTGCTTCTGCTTCTGCTGCAATGATCTCGAATCTTCAAC 2373
 QY 2394 GGAATTCAGCAGGAGGAGTGTGCTTCTGCTGCAATGATCTCGAATCTTCAAC 2453
 Db 2374 GGAATTCAGCAGGAGGAGTGTGCTTCTGCTGCAATGATCTCGAATCTTCAAC 2433
 QY 2454 GCTTACGACGAGTGTGCAACCGGTCACATTCGCTGCAATGATCTCGAATCTTCAAC 2513
 Db 2434 GCTTACGACGAGTGTGCAACCGGTCACATTCGCTGCAATGATCTCGAATCTTCAAC 2493
 QY 2514 GCGAGTTCAGGAGGAGTGTGCAACCGGTCACATTCGCTGCAATGATCTCGAATCTTCAAC 2573
 Db 2494 GCGAGTTCAGGAGGAGTGTGCAACCGGTCACATTCGCTGCAATGATCTCGAATCTTCAAC 2553
 QY 2574 GTCTTCAGGCAAGAGCGGCTTCTGCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2633
 Db 2554 GTCTTCAGGCAAGAGCGGCTTCTGCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2613
 QY 2634 AGGCTCTGCAAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2693
 Db 2614 AGGCTCTGCAAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2673
 QY 2694 GGCATCGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2753
 Db 2674 GGCATCGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2733
 QY 2754 GACTCAGTCACTGCTCCGATCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2813
 Db 2734 GACTCAGTCACTGCTCCGATCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2793
 QY 2814 GTGGAGCTCAATCAATGCTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2873
 Db 2794 GTGGAGCTCAATCAATGCTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2853
 QY 2874 GCGCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2933
 Db 2854 GCGCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2913
 QY 2934 AGATGATCAACCAAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2993
 Db 2914 AGATGATCAACCAAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2973
 QY 2994 AAGCAATCTCAAGTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3053
 Db 2974 AAGCAATCTCAAGTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3030
 QY 3054 ACCCCCTGGGAGTATTTGATTTCAACAGATTTCAACGATTTCAACGATTTCAACGAT 3113
 Db 3031 ACCCCCTGGGAGTATTTGATTTCAACAGATTTCAACGATTTCAACGATTTCAACGAT 3090
 QY 3114 CAGGATCTATCAACCAATTTGGGATTTCCGCGCCCAAGAGCTCAACTTCAACTCTT 3173
 Db 3091 CAGGATCTATCAACCAATTTGGGATTTCCGCGCGCAAGAGCTCAACTTCAACTCTT 3150
 QY 3174 AACATCAAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGG 3233
 Db 3151 AACATCAAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGG 3210
 QY 3234 ACCAGCAGGTTCAAGTCTTCTGCACTGAGTCAAGGAGTCAAGGAGTCAAGGAGTCA 3293
 Db 3211 ACCAGCAGGTTCAAGTCTTCTGCACTGAGTCAAGGAGTCAAGGAGTCAAGGAGTCA 3270
 QY 3294 GCGCAGCAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3353
 Db 3271 GCGCAGCAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3330

OY	3354	TACCTGACGCTCAACAATGGGAGCCAAACCGTGGGAGCGTTCACTCTTTACTGCTGGAA	3413
Db	3331	TACCTCAACCTCGAACAACGGGAGTCAGGACAGTAGGAGCGCTTTCAATTTTACTGCTGGAG	3390
OY	3414	TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTATCTTACGTTACACCTTTGAG	3473
Db	3391	TACTTTCCTTCTCAGATCTGCTGATACCGGAACAACCTTATCTTACGTTACACTTTTGAG	3450
OY	3474	GAAATGCTTTTCCACAGAGCTTAGCCGACACGCCAGACCGCTGAGACCGCTGATTAATCT	3533
Db	3451	GACGTTCTTTTCCACAGAGCTTAGCCTACACGCCAGACGCTGAGACCGCTCATTAATCT	3510
OY	3534	CTCATCGACCAATACCTGTATTTACTGTAACAGAACTCAAAATTCAGTCGCGAAGTCCCAA	3593
Db	3511	CTCATCGACCAATACCTGTATTTACTGTAAGCAAAACACTCCAAAGTGGAAACCAACAG	3570
OY	3594	AACAAGACCTTGCTGTTTAAAGCCGAGGAGTCTCAAGCTGCAATGTCTGTTACGCCAAAAC	3653
Db	3571	CAGTCMAAGCTTCAAGTTTCTCAGGCCCGAAGCAAGTGCATTTCGGGACCAAGCTTAGAAC	3630
OY	3654	TGGCTACCTGGACCCCTGTATTCGGAGAGAGCGCGTTTATAAACAACAAACAGACAACAC	3713
Db	3631	TGGCTTCTGGACCTGTATTCGGACAGACGAGATTAAGAACAATCTGCGGATTAACAAC	3690
OY	3714	AACAGCAATTTTACCTGAGCTGTGCTTCAAAATTAACCTCAATGGCGTGAATTCATC	3773
Db	3691	AACAGTGAATCTCGTGGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG	3750
OY	3774	ATCAACCTGTGCACTGCTATATGGCCCTCAACAAGAAGCGAAGAACAGTCTTTCCATG	3833
Db	3751	GTGAATCCGGGAGCCCGCCATGGCAAGCCACAAGGACGTAGAGAAAGTTTTCCTCAG	3810
OY	3834	AGCGGTGCATGATTTTGGAAAGAGAGCGCCGAGCTTCAACACTGCATTGGAACAT	3893
Db	3811	AGCGGGGTTCATCTTTGGGAGCAAGGCTCAAGAAAACAAATGTGAACATTGAAAAG	3870
OY	3894	GTCATGATTTACAGACGAAAGAGAAATTAAAGCACTAACCTGTGGCCACCGAAGATTT	3953
Db	3871	GTCAATGATTTACAGAGAAAGAGAAATCGAAACAAACAAATCCGTGGCTACGAGACAGAT	3930
OY	3954	GGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACAGACCCGTGGACCCGAGATGTGCAT	4013
Db	3931	GGTTCGTATCTACCAACCTCCAGAGAGGACAAGACAAGAGGCTAACCGCAGATGTCAAC	3990
OY	4014	GCTATGGAGCATTAACCTGGGACGTGTGGCAGATAGAGAGTACCTGACGGGTCCC	4073
Db	3991	ACACAAAGCGTTCTTCCAGAGCATGTGTGGCAGGACAAGATGTGTACTTCCAGGGGCC	4050
OY	4074	ATTGGGGCAAAATTCCTCAACAAGATGGAACATTTACCCGTCCTCTTATGAGGAGGC	4133
Db	4051	ATCTGGCAAAAGATTTCAACAACGAGCGGACATTTTCAACCCCTCTCCCTCATGGGTGGA	4110
OY	4134	TTTGGACTCAAGAACCCGCGCTTCTCAGATCTCATCAAAAACAGCGCTGTTCTTGGCAT	4193
Db	4111	TTGCGACTTTAAACACCTCTCTTCAACAATTCATCAAGAAACACCCCGGTACTTGGCAT	4170
OY	4194	CCTCGCGGAGTTTCAAGCTACAAAGTTTGCTTCAATCATCAACCAATACCTCACAGGA	4253
Db	4171	CTTTCGACCACTTCAAGTCGGCAAAAGTTTGCTTCTTATCAACAATACCTCACAGGA	4230
OY	4254	CA-AGTGAAGTGAATAATGGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAAATCC	4312
Db	4231	CACGCTCAGCGTGGAGATCGATGGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAAATCC	4290
OY	4313	CGAATGCAATACATCAATTAATGCAAAATCTGCCAAGCTGATTTTACTGTGGACA	4372
Db	4291	CGAAATTCAGTACCTTCCAACTAACAAAGCTCTTATCTGTGACTTAACGTGGATAC	4350
OY	4373	CAATGGAATTTATACGAGCGCTGGCCCATTTGGACCCGTTACCTTACCGGCTCCGTTA	4432
Db	4351	TAAATGCGGTATTAAGAGCCCTGGCCCATTTGGACCCAGATTAAGCTGATCTGTAATCTGTA	4410

OY		4433	ATTACGCTTAATCAATAAAACCGGTGAATTCGTTAGTGAATCTTGCTCCTGCCT	4492
Dd		4411	ATTTCCTGTTAATCAATAAACCGGTGAATTCGTTAGTGAATCTTGCTCCTGCCT	4470
OY		4493	TCTTATCTTAATCCGGTTACCATAGTTAATAGCTTACACATTAAGTCTTGTTGCCCTTCGC	4552
Dd		4471	TTCCTTCTTAATCTAGTTTCCATGCTACGTAGAATTAATTAAGATGCGCGGTTAATCATTTAA	4530
OY		4553	GATTAAGACTTACGTCATCGGGTTACCCTTAGTATGAGATTGCCCACTCCCTCTCTGC	4612
Dd		4531	CTACAAGGA-----ACCCCTAGTGAATGAGATTGGCCAATCCCTCTCTCTGC	4574
OY		4613	GCGTCGCTGCTCGTGGTGGGGCTGCGGACCAACAAAGTCCGACAGCGCAGAGCTGTGCTC	4672
Dd		4575	GCGCTCGCTCGCTACTGAGAGCGCGGGGACCAAAAGTTCGCCGACGCCCGGGCTTTGCC	4634
OY		4673	TGCCGGCCCCACCGAGCGAGCGAGCGGCCGACAGAGGAGTGGGCAA	4718
Dd		4635	GGGCGGCGCTCAGTGAGCGAGCGAGCGCGCACAGAGGAGTGGCCAA	4680
<hr/>				
RESULT 8				
AH41481				
ID	AH41481	standard; DNA;	4675 BP.	
XX	AH41481;			
DT	23-AUG-2001	(first entry)		
DE	Adeno-associated virus (AAV) plasmid pAV1 Avari nucleotide sequence.			
KW	Adeno-associated virus; AAV; gene transfer; adenovirus; chromosome 19;			
KM	medicine; cell engineering; genetic engineering; treatment;			
KW	embryological engineering; ds.			
OS	Adeno-associated virus.			
PN	WO200132899-A1.			
PD	10-MAY-2001.			
PF	23-OCT-2000; 2000WO-JP07373.			
PR	29-OCT-1999; 99JP-0308839.			
PA	(TAKI) TAKARA SHUZO CO LTD.			
PI	Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K;			
DR	Asada K, Kato I;			
XX	WPI; 2001-316450/33.			
PT	Gene transfer method for highly efficient introduction of foreign genes			
PT	to cells in humans particularly by integration specifically into AAVSI			
PT	site of chromosome 19 for transformation, applicable e.g. in disease			
PT	treatment -			
PS	Example 1; Page 28-31; 38pp; Japanese.			
XX	The present invention describes a gene transfer method for highly			
CC	efficient introduction of foreign genes to cells in humans particularly			
CC	by integration specifically into AAVSI site of chromosome 19 for			
CC	transformation, applicable e.g. in disease treatment. The method			
CC	comprises transferring into the cells (with the use of an adenovirus			
CC	vector) a nucleic acid which has a sequence provided with adeno-			
CC	associated virus (AAV)-originated inverted terminal repeats (ITRs) in			
CC	both sides of the target foreign gene to be transferred, a second nucleic			
CC	acid which has an AAV-originated rep gene and a promoter for expressing			
CC	this gene and carries a stuffer sequence inserted into it sandwiched in			
CC	two recombinase-recognition sequences and located between the rep gene			
CC	and promoter and expressing the Rep protein under the action of the			
CC	recombinase in the cells obtained previously to integrate the target			
CC	sequence into the genome of the cells. The method is for transferring a			

CC foreign gene into cells particularly in human, especially by integration
CC specifically into AAVS1 site of chromosome 19 for transformation, which
CC is applicable in medicine, as well as cell, genetic and embryological
CC engineering e.g. in disease treatment. The method is highly efficient,
CC with use of adenovirus vectors, nucleic acids and other sequences
CC including rep genes and promoters, by expressing rep proteins to
CC integrate target foreign gene into chromosomal DNA through action of
CC recombinase. The present sequence represents an AAV plasmid pAV1 Aavl1
CC nucleotide sequence which is used in an example from the present
CC invention.

XX Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 64.8%; Score 3055.8; DB 22; Length 4675;

Best Local Similarity 79.9%; Pred. No. 0; Mismatches 887; Indels 61; Gaps 13;

Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCCACCTCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTTGGCGAACCAAGTCCG 60
DB 1 TTGGCCACTCTCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTTGGCGAACCAAGTCCG 60
QY 61 AGACGCGAGCTCTGCTCTGCGCGCTCGCTCGGTGGGCTTGGCGAACCAAGTCCG 120
DB 61 CGACGCGCGGCTTGGCGCGCTCGCTCGGTGGGCTTGGCGAACCAAGTCCG 120
QY 121 GGCACTCCATCTAGGGGTAAATCGGGAAGCGCTCCCAAGCTGCGCGCTCA 180
DB 121 GGCACTCCATCTAGGGGTAAATCGGGAAGCGCTCCCAAGCTGCGCGCTCA 180
QY 181 CGTAATTTAGTCAATAGG---GAGTGTCTGTATTAAGTCACTGAGTCTTTG 237
DB 164 CGTAATTTAGTCAATAGG---GAGTGTCTGTATTAAGTCACTGAGTCTTTG 237
QY 238 GACATTTTGGGCAACGAGTGGCCATTATTAATATGAGTGGAGTCAAGAT 297
DB 223 GACATTTTGGGCAACGAGTGGCCATTATTAATATGAGTGGAGTCAAGAT 297
QY 298 CTTCATTTTG---ACCGGCAATTTTGAAGAGAGAGAGAGAGAGAGT 356
DB 283 CTTCATTTTGAAGGAGGAGTGGTGAAGAGAGAGAGAGAGAGAGT 356
QY 357 TGATCAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 416
DB 343 TGATCAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 416
QY 417 GCTGGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 476
DB 403 ACTGGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 476
QY 477 TTGAGT 536
DB 463 TTGAGT 536
QY 537 GCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 596
DB 523 GCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 596
QY 597 ACTTTCACCTTCATATTCGTGGAGAGAGAGAGAGAGAGAGAGAGT 656
DB 583 ACTTTCACCTTCATATTCGTGGAGAGAGAGAGAGAGAGAGAGAGT 656
QY 657 TCTGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 716
DB 643 TCTGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 716
QY 717 TGCCCACTGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 776
DB 703 TGCCCACTGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 776
QY 777 TGAGAGAGTCAATCCCACTAATCTCTGCGCAAGTCAAGTCAAGT 836
DB 763 TGAGAGAGTCAATCCCACTAATCTCTGCGCAAGTCAAGTCAAGT 836

QY 837 CGTGAATCAATGAGAGAGATATTAAGAGAGAGAGAGAGAGAGAGAGT 896
DB 823 CGTGAATCAATGAGAGAGATATTAAGAGAGAGAGAGAGAGAGAGAGT 896
QY 897 TCGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 956
DB 883 TCGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 956
QY 957 ACCCAATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1016
DB 943 ACCCAATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1016
QY 1017 TCGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1076
DB 1003 TCGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1076
QY 1077 CTTGTAATCTCTCTTCAACGCGCTTCCAACTGCGGCTTCCAGATCAAGGCGCT 1136
DB 1063 CTTGTAATCTCTCTTCAACGCGCTTCCAACTGCGGCTTCCAGATCAAGGCGCT 1136
QY 1137 ACAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1196
DB 1123 ACAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1196
QY 1197 CTTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1256
DB 1183 CTTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1256
QY 1257 AACCTGCTCAACGCGGCTCTCTCTCTGCGGCTGCGGAGAGAGAGAGT 1316
DB 1243 AACCTGCTCAACGCGGCTCTCTCTCTGCGGCTGCGGAGAGAGAGT 1316
QY 1317 ACAATCTGAGTCTGTTGGGCGGAGAGAGAGAGAGAGAGAGAGAGT 1376
DB 1303 ACAATCTGAGTCTGTTGGGCGGAGAGAGAGAGAGAGAGAGAGAGT 1376
QY 1377 CCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1436
DB 1363 CCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1436
QY 1437 ATTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1496
DB 1423 ATTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1496
QY 1497 AGTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1556
DB 1483 AGTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1556
QY 1557 CCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1616
DB 1543 CCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1616
QY 1617 TTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1676
DB 1603 TTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1676
QY 1677 TTGACCTCAACGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGT 1736
DB 1663 TTGACCTCAACGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGT 1736
QY 1737 AGTTCTTCGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1796
DB 1723 AGTTCTTCGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1796
QY 1797 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1856
DB 1783 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1856
QY 1857 CTTGCGGCTCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1916
DB 1843 CTTGCGGCTCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1916
QY 1917 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1976

Db	1900	CAGAGAGGATACCAAAACAAATGTTCTGCTACGAGGAGATGATCTGATGCTGTTTCCCT	1959
OY	1977	GCAGACATGCGAGAGATGAAATCAGAATTTCAACATTTGCTTCACGACGGAGCGAG	2036
Db	1960	GCAGACATGCGAGAGATGAAATCAGAATTTCAAAATCTGCTTCACTCACGAGACGAAG	2019
OY	2037	ACTGTTCAAGATGCTTCCCGGCGTGCAGAACTCAACCGGT--CGTCAGAAAGAGGA	2093
Db	2020	ACTGTTTAGAGTGCTTTCC--CGTGCAGAAATCAACCGTTCGTGCTCAAAAAGG	2076
OY	2094	CGTATCCGAAACTCTGTGCAATTCATCTGTCTGGGGCGGACTCCGAGATGCTTGCT	2153
Db	2077	CGTACCAAGAACTGTGCTACATTCATTCATCATGGAAGATGACAA--CGCTTGCA	2133
OY	2154	CGGCTCGGATCTGATCAACGTCGACCTGATGATCTGTGTTCTGAGCAATTAATGACTT	2213
Db	2134	CTGCTCGGATCTGATCAATGTGGAATTTGATGATCTGCATCTTTGAACAAATTAATGATTT	2193
OY	2214	AAACAGGATATGCTGCCGATGTTATCTTCCAGATTTGGCTTCAGAGCAACCTCTGTAG	2273
Db	2194	AAATCAGGATATGCTGCCGATGTTATCTTCCAGATTTGGCTTCAGAGCAACTCTGTGAA	2253
OY	2274	GGCATTCGCGATGCTGTGGGACTTGAACCTTGAGCCCGGAAGCCCAAGCCACCAAGAA	2333
Db	2254	GGAATTAACAGATGTGTGGAAGCTCAACCTGCCCCACCAACCAAAACCCGACAGAGCG	2313
OY	2334	AAGCAGACAGACGCGCGGGGATCTGTGCTTCCGTGCTCAATATACCTGACCCCTTCAAC	2393
Db	2314	CATAAGACAGACAGAGGGGCTTGTGCTTCTGTGGATCAATATACCTGACCCCTTCAAC	2373
OY	2394	GGACTCGACAAAGGGGAGCCCGTCAACGCGCGAGCGAGCGAGCCGCTTCGACACG-ACAA	2452
Db	2374	GGACTCGACAAAGGGGAGCCCGTCAACGAGGACAGCGCGAGCCGCTTCGACACGTAACAA	2433
OY	2453	GGCTTACGACCGACGACTTCAAAAGCGGATGACAAATCCGATCTGTGGGATTAACACGCCGA	2512
Db	2434	AGCTTACGACCGGACGCTCGACAGCGGAGACAAACCGTATCTCAATGACAAACAGCCGA	2493
OY	2513	CGCGAGTTTCAGAGCGCTGTCAGAAAGATACGCTTTTGAGGCAACCTCGGAGCAGC	2572
Db	2494	CGCGAGTTTCAGAGCGCCTTTAAAGAAATACGCTTTTGAGGCAACCTCGAGCAGC	2553
OY	2573	AGTCTTCAGGCGCAAGAGCGGGTCTCGAACTCTCGGTGTGTTGAGAGGCGCTAA	2632
Db	2554	AGTCTTCAGGCGCAAGAGGGGTTCTTGAACTCTTGAGCCGTGTTGAGAACTGTTTAA	2613
OY	2633	GACGCTCTTGAAAAGAAACGTCCGGTGAAGATGCGCACAAAGCCAGACTCTCTCTC	2692
Db	2614	GACGCTCTCGGAAAAAAAGAGGCGCGTGAAGACTCTCTGTGAGCCAGACTCTCTCTC	2673
OY	2693	GGGCACTCGGCAAGACAGGCGACGACCCGCTAAAAAGACATCAATTTTGGTCAACTGG	2752
Db	2674	GGGAACCGGAAGAGCGGGCGACGACCTGCAAGAAAAAGATTTGATTTGGTCAACTGG	2733
OY	2753	CGACTCAGATGACGTCCCGATTCACAACTCTCGAGAACCTTCAGCAACCCCGCTGC	2812
Db	2734	AGACGCAAGACTCAGTACCTGAACTCCCAAGCTCTCGGACAGCAACAGGACGCCCTCTGG	2793
OY	2813	TGTGGGACTCTACTAATAGGCTTCAGGCGGTGAGCCACCAATGTGAGACATTAACGAAG	2872
Db	2794	TCTGGGAATCTAATAGATGGCTACAGGAGTGGCCCAACCAATGTGAGCAATTAACGAAG	2853
OY	2873	CGCGACGGAAGTGGTAAATGCTTCAGAAATTGGATTGGCATTCACATGAGCTGGGCGA	2932
Db	2854	CGCGACGGAAGTGGTAAATTTCTCCGGAATTTGGATTTGCAATTTGCAATGATGGGCGA	2913
OY	2933	CAGATCTACACACACAGACCCGACCTGGGCTTGGCCACTTACATTAACCACTCTA	2992
Db	2914	CAGATCTACACACACAGACCCGAACTGTGGCCCTGGCCACTTACATTAACCACTCTA	2973
OY	2993	CAGCAATCTTCAGTGTCTCAACGGGGCCAGCAACGACACATCTATCTTGGCTACAG	3052

Db	2974	CAAAACAATTTCCAGCCAATATA---GGAAGCTCGAAGCAATATCACTATCTTGGCTACAG	3030
QY	3053	CACCCCTCGGGGGATATTTGATTTCAACAGATTCACATGCTCCACTTTTCACCAAGCTGAC	3112
Db	3091	CACCCCTTTGGGGGATATTTTGACTTCAACAGATTCACATGCTCCACTTTTCACCAAGCTGAC	3090
QY	3113	GGAGGACTCATCAACAACAAATTGGGGATTTCCGGCCCAAGAGACTCAACTTCAAACTCTT	3172
Db	3091	GCAAAAGACTCATCAACAACAACCTGGGGATTCGAGCCCAAGAGACTCAAACTTCAAGCTCTT	3150
QY	3173	CAACATCCAAATCAAGGAGGTCAAGAGCAATATGGCGGTCAACAACATCGCTAATTAACCT	3232
Db	3151	TAACTTTAAATCAAAAGAGTCAAGCAGAAATGAACGTATCAGACAGATTTGCCAATTAACCT	3210
QY	3233	TACGAGCAGGTTCAAGTCTTCTCGAATCTCGAATCAACAGTTCGATAGCTTCGAGCTC	3292
Db	3211	TACGAGCAGGTTCAAGGTTGATTTACTGAATCGGAATGACAGCTCCGATACGTCCTCGAGCTC	3270
QY	3293	TGCGCACAGGGCTGAGCTCTCCTCCGTTTCCGGCGGAGCTGTTCAATGATTTCCGCAATACGG	3352
Db	3271	GAGCGATCAAGAAATGCGCTCCGCGGTTCCGACAGACGTCTTCAATGGGCCACAGATATGG	3330
QY	3353	CTACCTGACGCTCAACAATGGCAGCGAAGCGGTGGAGAGGTCACTCCTTTTATCTGCGCTGGA	3412
Db	3331	ATACCTCAACCTGAACAACGGAGTCAAGCATGTAGAGAGCTCTTCAATTTTATCTGCGCTGGA	3390
QY	3413	ATATTTCCCTTCTGAGATGCTGAGAACGGGCAACAATTATCTTCACTTCACTACACTTTTGA	3472
Db	3391	GTACTTTCCTTCTGAGATGCTGCGAATCCGAAACAATTATCTTCACTTCACTACACTTTTGA	3450
QY	3473	GGAAGTGCCTTTCCACAGAGACTACGGCAACGCCAAGAGCTGGACCGGCTGATGATATCC	3532
Db	3451	GGAAGTTCCTTTCCACAGAGACTACGCTCAAGCGAGAGTGGACCGTCACTGAATATCC	3510
QY	3533	TCTCATGAGCAATACCTGTATTACTGTAAACAGAACTCAAAATCAAGTCCGGAAGTGCCTCA	3592
Db	3511	TCTCATGAGCAATACCTGTATTACTGTAAACAGAACTCAAAATCAAGTGAACCACTAC	3570
QY	3593	AAACAAGACTTGTGTTTAAACCGTGGGTCTCCAGCTGGCAATGTCGTTCAGCCCAAAA	3652
Db	3571	GCAATCAAGCTTCAAGTTTCTCAGAGCCGAGCGAGATGACATTTGGGGACAGTCTAAGAA	3630
QY	3653	CTGGCTACTGGAGCCCTGTTATCGGACAGAGCGGCTTCTTAAACAAAAACAGACACAA	3712
Db	3631	CTGGCTTCTGGAGCCCTGTTATCCGACAGACGAGTATCAAAAGCATCTGGGATTAACAA	3690
QY	3713	CAACAGCAATTTTAACTGGAATGGTGTCTTAAATTAACCTCAATGGGCGGTGAATCAT	3772
Db	3691	CAACAGTAAATCTGTGGACTGGAGCTCAACAGTACCACTCAATGGCAGAGACTCTCT	3750
QY	3773	CATCAACCTTGACATGCTATAGGCTCAACAAGACGACGAAACAAGTTCTTTTCCAT	3832
Db	3751	GGTGAATC-----CGGCAATGGGCAAGGCAACAGAGATGAABAAAAAGTTTTTCTCA	3804
QY	3833	GAGCGGTGTCAATGATTTTGGAAAAAGAGCGCCGAGACTTCAAAACACTGATTTGGACAA	3892
Db	3805	GAGCGGGGTTCTCACTTTGGGAAAGCAAGCTCTCAGAGAAAAAATATGTGAACATTTGAAA	3864
QY	3893	TGTATGATTTCAACAGCAAGAGAAATTAAGCCACTAACCCTGGGCAACGAAAGATT	3952
Db	3865	GGTCAATGATTTCAACAGCAAGAGAAATCGAACAACCAATCCGTTGGCTACGAGACAGTA	3924
QY	3953	TGGACACGTGACATTAATTTCCAGAGCAGCAGACAGACCTTCGACCGGAGATGTGCA	4012
Db	3925	TGTTTCTGTATCTAACAACTCTCCAGAGAGCGCAACAGAACACTACCGGAGATGTCAA	3984
QY	4013	TGCTATGGAGCAATTAACCTGGACATGTGTGCAAGATAGAGAGCTGTACTTCAGAGGCTC	4072
Db	3985	CACACAAGGCGTTCTTCCAGGCAATGCTCTGGCAAGACAGAGATGTGTAACCTTCAGGGGCC	4044
QY	4073	CATTTGGGCAAAATTTCTCTCAACAGATGACATTTTCAACCCGTCCTCTTATGGGCGG	4132
Db	4045	CATCTGGGCAAAAGTTCCACACAGAGCAGACATTTTCAACCCCTCTCTCTCATAGGGGTGG	4104

QY 4133 CTTTGAGCTCAAGAACCCGCTCTCAGATCTCATCAAAAACAGCGCTTCTCGGAA 4192
 DB 4105 ATTGCGACTTAAACACCCCTCTCCACAGATTCATCAAGAAACACCCCGTACTCGAA 4164
 QY 4193 TCTCCGCGCGAGTTTACGCTCAAAAGTTGGCTTCATTCATCAACCCCAATCTCAAG 4252
 DB 4165 TCTTCGACCACTTCAGTGCAGCAAGTTGGCTTCCTTCATCAACAGTACTCCAGGG 4224
 QY 4253 ACA-AGTGAATGGAATTTGATTTGAGAGCTGCAGAAAGAAACAGCAAGCTTGAAATC 4311
 DB 4225 ACACGGTCAACCGTGAATGAGTGGAGCTGCAGAAAGAAACAGCAAGCTTGAAATC 4284
 QY 4312 CCGAAGTCACTACATCCCAATTTATGCAAAATCTGCAAGCTTGAATTTACTGTGACA 4371
 DB 4285 CCGAAATTCAGTACATTCCTCACTCAACAGTCTGTAATCGTGAATTTACTGTGACA 4344
 QY 4372 ACAATGACTTTTACTAGAGCTTGCAGCCCAATGGAACCCGTTACCTTACCCGTCCTGT 4431
 DB 4345 CTAATGGCGGTATTCAGAGCTTGCAGCCCAATGGAACCCGTTACCTTACCCGTCCTGT 4404
 QY 4432 AATTAGCTGTTATTCATTAACCGGTTGATTCGTTTCACTTGAATTTGTCTCTCTCC 4491
 DB 4405 AATTGCTTGTATTCATTAACCGGTTGATTCGTTTCACTTGAATTTGTCTCTCTCC 4464
 QY 4492 TTCTTATCTATCGTTACCATGATGTTATAGCTTACACATTAACCTGCTTGTGCGCTCG 4551
 DB 4465 TTCTTATCTATCGTTACCATGATGTTATAGCTTACACATTAACCTGCTTGTGCGCTCG 4524
 QY 4552 CGATTAAGACTTACGTCATCGGTTACCCCTAGTGAAGAGTTGCCACTCCCTCTCTG 4611
 DB 4525 ACTACAGAGA-----ACCCTAGTGAAGAGTTGCCACTCCCTCTCTG 4568
 QY 4612 CGGCTCGCTCGCTCGGTTGCGGCTGCGGACCAAAAGTCCGCAAGAGGAGACTCTGCT 4671
 DB 4569 CGGCTCGCTCGCTCGGTTGCGGCTGCGGACCAAAAGTCCGCAAGAGGAGACTCTGCT 4628
 QY 4672 CTGCGGCGCCCAACGAGCGAGCGCGCAGAGAGGAGTGGGCA 4718
 DB 4629 CGGCGCGCTCACTGAGCGAGCGCGCAGAGAGGAGTGGGCA 4675

RESULT 9
 AAF89931 standard; DNA; 4675 BP.
 AAF89931;
 06-AUG-2001 (first entry)

Nucleotide sequence of an adeno-associated virus Rep78 protein.
 adeno-associated virus; AAV; Rep78; papillomavirus-associated disease;
 cancer; HIV-associated disease; ss.
 Adeno associated virus.
 Key Location/Qualifiers
 CDS 321..2186
 /tag= a
 /product= "Rep78"

MO200132711-A2.
 10-MAY-2001.
 23-OCT-2000; 2000MO-US29187.
 21-OCT-1999; 99US-0160608.
 (UYAR-) UNITV ARKANSAS.
 Hermonat PL;

XX WPI: 2001-343480/36.
 DR P-PSDB; AAB83954.
 XX New adeno-associated virus AAV Rep78 mutant, useful for treating
 PT papillomavirus-associated diseases, cancer, or human immunodeficiency
 PT virus-associated diseases
 XX
 PS Disclosure; Fig 15; 53pp; English.
 XX
 CC The present sequence encodes an adeno-associated virus (AAV) Rep78
 CC protein. The specification describes an adeno-associated virus (AAV)
 CC Rep78 mutant, comprising an AAV Rep78 modified protein that possesses
 CC different biochemical and biological functions as compared to the
 CC wild-type AAV Rep78 protein. The AAV Rep78 mutant and associated DNA
 CC sequences can be used in a pharmaceutical composition. The pharmaceutical
 CC composition can be used for treating papillomavirus-associated diseases,
 CC cancer, or HIV-associated diseases.
 CC
 SQ Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 64.8%; Score 3055.8; DB 22; Length 4675;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGGCCACTCCCTCTCTGCGGCTGCTGCTCGGTTGGGCGCTGCGGACCAAGTCCGC 60
 DB 1 TTGGCCACTCCCTCTCTGCGGCTGCTGCTCGGTTGGGCGCTGCGGACCAAGTCCGC 60
 QY 61 AGACGCGCAGAGCTCTGCTCTGCGGCGCCACGAGCGAGCGCGCAGAGGAGATG 120
 DB 61 CGACGCGCGGCTTGGCGGCGGCGCTGAGAGAGAGCGAGCGGAGGAGATG 120
 QY 121 GGCACCTCATCACTAGGAGTAAATGCGAAGCGCTTCCACGCTGCGGCTCAAGCCTGA 180
 DB 121 GCCAATCCATCACTAGGAGTAAATGCGAAGCGCTTCCACGCTGCGGCTCAAGCCTGA 163
 QY 181 CGTAATTCAGTCACTAGGAGG---GAGTGTCTGTATTTAGCTGACAGTGAAGTCTTTGC 237
 DB 164 CGTAATTCAGTCACTAGGAGGTTTGAAGAGTCTGTATTTAGTCACTAGTGAAGTCTTTGC 222
 QY 238 GACATTTTGCAGACACGCTGCGCATTTAGGATATATATGCGCGAGTGAAGCAGAGAT 297
 DB 223 GACATTTTGCAGACACGATGCTGCAAGCTGGATATTTAGCGAGTGAAGCAGAGAT 282
 QY 298 CTCATTTTGG-ACCGGAAATTTGAACGAGCAGAGCCTGCGGCTTTACAGAGATG 356
 DB 283 CTCATTTTGAAGCGGAGGATTTGAACGAGCAGAGCCTGCGGCTTTTACAGAGATG 342
 QY 357 TGATCAAGGTGCGAGCGACCTGAGCGACCTGCGGCGCATTTCTGACTGCTTGTGA 416
 DB 343 TGATTAAGGTGCGAGCGACCTTGAAGCGGATCTGCGGCGCATTTCTGAAGTCTTGTGA 402
 QY 417 GCTGGGTGCGCAGAGAGATGGAGCTGCCCCGAGATTCTGACATGATCTGAATCTGA 476
 DB 403 ACTGGGTGCGCAGAGAGATGGAGTGGCCGAGATTCTGACATGATCTGAATCTGA 462
 QY 477 TTGAGCAGGCACTCCCTGACCTGCGCGGAGAGCTGAGCGGCACTTCTGCTCAATGCG 536
 DB 463 TTGAGCAGGCACTCCCTGACCTGCGCGGAGAGAGCTGAGCGGCACTTCTGAGAGATGCG 522
 QY 537 GCGGCGTGAATGAGGCGCGGAGGCGCTTCTTTGTTAGTTGAGTGAAGGCGAGTCT 596
 DB 523 GCGGCGTGAATGAGGCGCGGAGGCGCTTCTTTGTTAGTTGAGTGAAGGCGAGTCT 582
 QY 597 ACTTCACCTCCATATTTGTTGAGAGACCGGCGGTCAATTCATGATGCTGCGGCGCT 656
 DB 583 ACTTCACATGACAGTGTGTTGAACACCGGCGGTGAATTCATGATGTTTGGGCGTT 642
 QY 657 TCTGAGTGAATTTAGGAGCAAGCTGTGAGACCATCTACCGGCGGATGAGCGGACCC 716
 DB 643 TCTGAGTGAATTTGCGGAAAACTGATTCAGAGAAATTTACCGGCGGATGAGCGGACTT 702

OY	717	TGCCAATGGATTGGCGGTGACCAAGAAGCGCTAATGGCGCCGAGGGGGGAAACAAGTGG	776
Db	703	TGCCAATCGTTTGGCGGTCACAANAGACCAAAATTGGCGCCGAGGGGGAAACAAGTGG	762
OY	777	TGAGCAGATGGTACATCCCACTACTCTCTGCCAAGACTGACGCCGAGCTGACGTGG	836
Db	763	TGATATGATGTCTACATCCCCCAATTACTTGCTCCCCAAACCACACTGAGGTCTCAGTGG	822
OY	837	CGTGAACCTAATAGTAGAGATATATATAGCGCTGTGTTTGAACCTTGCGGAGGCAACGGC	896
Db	823	CGTGACCTAATATAGAACAGTATTTTAAAGCGCTCTTTTGAATCTCAAGAGGCTTAAACGT	882
OY	897	TCTGGCGCAGACACCTGACCCACGTCAGCCAGACCCAGAGTAGAACAGAGATCTGA	956
Db	883	TGTTGGCGCAGACATCTGACGACCGGTGTGACAGACGACAGAGACAAACAAGAGATCAGA	942
OY	957	ACCCCAATTCTGACGCGCGCTGTCAATCCGGTCAAAAACCTCCGCGCGCTACATGGAGCTGG	1016
Db	943	ATCCCAATTCTGATGCGCGCGGTGATCAATCAAAAATTCAAGCAGGTACTGAGAGCTGG	1002
OY	1017	TCGGGTGCTGATGAGCCGGGGCATCATCTCCAGAGACAGTGGATCCAGAGAGACCAAG	1076
Db	1003	TCGGGTGCTGTGAGCAAGGGGATTACTCGAGAGACAGATGGATCCAGAGAGACCAAG	1062
OY	1077	CCTGTACATCTCTTCAACGCGCTTCAACTGCGGTCCCAATCAAGCGCTCTGG	1136
Db	1063	CCTCATATCTCTCTCAATGCGGCTCAACTGCGGTCTCCAAATCAAGGCTCTGG	1122
OY	1137	ACAATGCGGAGAGATCAATGGCGCTGACCAATCCGCGCCGACATCACTGGTATGAGCCCG	1196
Db	1123	ACAATGCGGAGAGATTAATGAGCTGACTTAAACCGCCCGACTACTCTGTTGGGCCAGC	1182
OY	1197	CTCGCCCGCGAGCATTTAAACCAACGCGCATCTACCGCATCTGAGCTGACGCGTACG	1256
Db	1183	AGCCCGTGAAGACATTTCCAGCATCGGATTTTAAATTTGGAATTAACCGGTACG	1242
OY	1257	AACTGCGCTAAGCGCGCTCCGCTTCTCCGCTGGGCGCCAGAAAAGTTGGGAAAGCGCA	1316
Db	1243	ATCCCAATATGCGGCTCTCCGCTTCTTGGAGATGGGCCAGAAAAGTTCCGCAAGAGGA	1302
OY	1317	ACACCATCTGGCTGTTTGGGCGCGGACCAACGCGGACAGACCAATCGCGGAAACCTCG	1376
Db	1303	ACACCATCTGGCTGTTTGGGCGCTGCAATCAACGCGGAAACCAATCGCGGAGCCATAG	1362
OY	1377	CCCAAGCCGTCCTTCTTAAGCGCTGCGCTCACTGACCAATGAAACTTTCCTTCATG	1436
Db	1363	CCCAACATGTGCTTCTTCAAGCGGTGCGTAACCTGACCAATGAAACTTTCCTTCACG	1422
OY	1437	ATTGCGTGCACAAATGATGATCTGTGTGGAGAGGACAAATGACGCGCAAGTCTGTG	1496
Db	1423	ACTGTGTGCACAAATGATGATCTGTGTGGAGAGGAGGAATGACCGCCCAAGTCTGTG	1482
OY	1497	AGTCGCGCAAGGCCATTTCTCGGCGGACGAAAGTGCCTGTGACCAAAAGTGCAGTCT	1556
Db	1483	AGTCGCGCAAAAGCCATTTCTCGGAGGAAACAAAGTGCCTGTGACCAAAAGTGCAGTCT	1542
OY	1557	CGGCGCAATGAGACCCGCAACCCCGGATGCTCACTCCAAACCAAACTGTGCGCGGTGA	1616
Db	1543	CGGCGCAATGAGACCCGCACTCCCGATGCTCACTCCAAACCAAACTGTGCGCGGTGA	1602
OY	1617	TTGACGGGAAACGACCACTTTCAGACACCAAGACCGTTGAGGACCGGATGTTCAAT	1676
Db	1603	TTGACGGGAAACTCAACGACCTTTCGAACACCAAGCACCGTTTCAGAGACCGGATGTTCAAT	1662
OY	1677	TTGAACTCAACCGCGCTGTGAGAGTACTTTGGAAAGTGAACAAGAGAGAAAGTCAAG	1736
Db	1663	TTGAACTCAACCGCGCTGTGATCACTTTGGAAAGTGAACAAGAGAGAAAGTCAAG	1722
OY	1737	AGTTCTTCCGCTGGCGCAGAGATCACTGACCGAGGTGGCGATGATGTTCTACGTCAGA	1796
Db	1723	ACTTTTTCGCGTGGGCAAAAGATCACTGTGTTGAGTGTGAAGATGAATTTCTACGTCAGA	1782
OY	1797	AGGGTGAACCAAAAGACCCGCGCCCGATGACCGGATTAAGACGACCCCAACGCGG	1856

Db	1783	AGGGTGAAGCCAAAGAAAGACC	GGCCCCCGAGTGA	CGCAGATATATTAAGGAAACCCCAAGGGG	1842	
QY	1857	CTTGCCCCCTCAGTCCGGATTCATTCGACGATCAGACGGGGAAGGAGCTCCGGTGGACCTTGG			1916	
Db	1843	TGCGCGAGTACGTTGGCCAGCGCATCGACGTCAGCGGGA	---	AGCTTCAATCAACTACG	1899	
QY	1917	CCGACAGGTACCAAAACAAATGTTCTCGTCA	CGCGGGCATGCTTCAGATCTGTTCCT		1976	
Db	1900	CAGACAGGTACCAAAACAAATGTTCTCGTCA	CGCGGGCATGATCTATCTGTTTCCT		1959	
QY	1977	GCACGACATCGGAGAGAAATGAATTCAGAAATTTCCAACTTTCCTTCA	CGCCACGGACGAGAG		2036	
Db	1960	GCACACAAATGGAGAGATGAATTCAGAAATTCAAATATCTCTTTCATCTCA	CGGACAGAAAG		2019	
QY	2037	ACTGTTACAGATGTTTCCCGGCGCTGCAGATCTCAACCGGT	---	CGTGAAGAAAGGA	2093	
Db	2020	ACTGTTTAGATGCTTTC	---	CGTGCAGATCTCAACCGTTC	2076	
QY	2094	CGTATCGGAAA	CTCTGTGCCAATTCATCATCTGCTGGGGCGGGCTCCGAGATTGCTTGTCT		2153	
Db	2077	CGTATCGAAGAA	CTGTGCTACATTCATCATATTCATGAGAAAGTGCACAGA	---	CGCTTGCA	2133
QY	2154	CGGCTCGCATCTGGGTCAACGGTGGACCTGGATGCTGTGTTTCTGACCAATTAATGACTT			2213	
Db	2134	CTGCTCGGATCTGTCTCAATGTGCAATTTGGATGATCTGCATCTTTGAACAAATTAATGATTT			2193	
QY	2214	AAACAGGATATGCTGCCGATGATGTTATCTTTCAGATTGGCTCGAGGACAA	CTCTCTGAG		2273	
Db	2194	AAATCAGGATATGCTGCCGATGATGTTATCTTTCAGATTGGCTCGAGGACAA	CTCTCTGAA		2253	
QY	2274	GGCATTCGGCAGTGTGTGGGACTTGA	AAACCTGGAGCCCGAAAGCCCAAGCCCAACACAGCA		2333	
Db	2254	GGAAATTAAGACATGTGTGAAGCTTA	AAACCTGGCCACCAACCAAGAGCCCGCAGAGCG		2313	
QY	2334	AAGCAGACGACGGCGGGGTCTGGGTCTTCTGGCTACAAATGACTCTCGACCTTCAAC			2393	
Db	2314	CATAAGGACGACACACAGGGGTCTTGTGCTTCTGGGTACAAATGACTCTCGACCTTCAAC			2373	
QY	2394	GGACTCGACAAAGGGGAGACCCGTCAACGCGCGGACGACGAGGGCCCTCGAGACAG	-ACAA		2452	
Db	2374	GGACTCGACAAAGGGGAGACCCGTCAACGAGGACACCCCGGGCTTCGAGCAAGTACAA			2433	
QY	2453	GGCTCTCGACACGAGACTCAAAAGGGGTGACAAATCCGTACTCTGGGTATACAAACGCGCA			2512	
Db	2434	AGCTTACGACCCGGAGCTCGACACCGGAGACAA	CCGTAACCTCAAGTACAAACACGCGCA		2493	
QY	2513	CGCGAGTTTCAGAGCGTCTGCAAGAGATATCCTCTTTTGGGGGCA	CTCGGGCGAGC		2572	
Db	2494	CGCGAGTTTCAGAGCGCTTAAAGAAATACCTCTTTTGGGGGCA	CTCGGACAGC		2553	
QY	2573	AGTCTTCACGGCCAAAGAGGGGTTTCGAA	CCCTCGGCTGTGTTGAGGAAGGCGCTAA		2632	
Db	2554	AGTCTTCACGGCCAAAGAGGGGTTCTTGA	ACCTCTGGGCTGTGAGGAACCTGTTAA		2613	
QY	2633	GACGCTCTTGGAAGAA	CGTCCGGTAGACAGTGCACAAAGCCGACCTCTCTCTC		2692	
Db	2614	GACGCTCTTGGAAGAAAGAGCGGTAGAGCA	CTCTCTGTGAGCCAGACTCTCTCTCTC		2673	
QY	2693	GGGCAATCGGCAAGACAGGCCAGAGCCCGCTTAA	AAAGATTCAAATTTGGTCAAGCTGG		2752	
Db	2674	GGGAACCGGAAGCGGGCCAGCGCTCGCAAGAA	AAAGATTTGGTCAAGCTGG		2733	
QY	2753	CGACTCAGAGTCAATCCCGATTCACAA	CTCTCGAGAACTTCAGCAACCCCGCTGC		2812	
Db	2734	AGAGCGAGACTCAATACCTGACCCCGACCT	CTTGGACAGCACACAGAGCCCTCTCTGG		2793	
QY	2813	TGTGGACCTTACTAATATGCTTCA	GGCGGTGCGCACCAATTGGCAGCAATAACGAAG		2872	
Db	2794	TCTGGAACTAATACATGTGCTTAC	AGGAGTGGCGCACCAATTGGCAGCAATAACGAAG		2853	
QY	2873	CGCGACGGAATGGGTATATGCTCAGGAA	ATTTGGATTCGATTCACATGGCTGGGGA		2932	

Db 2854 CGCGGACGAGTGGGTAAATTCCTCCGGAATTTGGCATTTGCGATTCCATGATGAGGCGA 2913
 QY 2933 CAGAGTCATCAACACGACACCCGACACTGGGCTTGGCCACCTACATTAACACCTCTA 2992
 Db 2914 CAGAGTCATCAACACGACACCCGACACTGGGCTTGGCCACCTACATTAACACCTCTA 2973
 QY 2993 CAGGCAATCTCGAGTCTTCAACGAGGAGCGACGACGACCAACCACTACTTGGCTACG 3052
 Db 2974 CAAGCAATTTCCAGCAATCA---GGAGCTGGAAGACATCACTACTTGGCTACG 3030
 QY 3053 CAGCCCTGGGAGTATTTGATTTCAACAGTTCCACTGCCACTTTTACCAAGCTGACTG 3112
 Db 3031 CAGCCCTGGGAGTATTTGATTTCAACAGTTCCACTGCCACTTTTACCAAGCTGACTG 3090
 QY 3113 GCGAGCTCATCAACAAATTGGGATTTCCGGGCGGAGAGTCAACTTCAAACTCT 3172
 Db 3091 GCGAGCTCATCAACAAATTGGGATTTCCGGGCGGAGAGTCAACTTCAAACTCT 3150
 QY 3173 CAACATCAAGTCAAGAGGTCAACGAGATGATGCGTCAACAACCTGCTAATACT 3232
 Db 3151 TAACATTCAGTCAAGAGGTCAACGAGATGATGCGTCAACAACCTGCTAATACT 3210
 QY 3233 TACCAAGACGCTTCAAGTCTTCTGAGCTGAGATCAAGCTTCCGTAAGCTCTGCTC 3292
 Db 3211 TACCAAGACGCTTCAAGTCTTCTGAGCTGAGATCAAGCTTCCGTAAGCTCTGCTC 3270
 QY 3293 TGGGACACAGGAGCTCCCTCCGTTCCGGGCGGAGAGTGTATGATTCGGAATACG 3352
 Db 3271 GGGGACATCAAGATGCTCCCGCGGTTCACAGACAGCTCTCAATGATGACAGATG 3330
 QY 3353 CTACCTGACGCTCAACAAATGAGCGACGAGCGTGGAGCTTCACTCTTACTGCTGGA 3412
 Db 3331 ATACCTACGCTGAACAGCGAGGTCAAGCATGAGACGCTCTTCAATTTACTGCTGGA 3390
 QY 3413 ATATTTCTCTTCTGAGTCTGAGAACGCGCAACACTTTACCTTCACTGCTGGA 3472
 Db 3391 GTATTTCTCTTCTGAGTCTGAGAACGCGCAACACTTTACCTTCACTGCTGGA 3450
 QY 3473 GGAAGTCTCTTCCACAGAGCTACGCGACAGCGACAGCTGGAACGGGCTGATGATTC 3532
 Db 3451 GGAAGTCTCTTCCACAGAGCTACGCGACAGCGACAGCTGGAACGGGCTGATGATTC 3510
 QY 3533 TCTCATGACCAATACCTGTAATTAAGTGAACAGAACTCAAAATCAGTCCGAGTCCCA 3592
 Db 3511 TCTCATGACCAATACCTGTAATTAAGTGAACAGAACTCAAAATCAGTCCGAGTCCCA 3570
 QY 3593 AAACAAAGACTTGTGTTTAAAGCGTGGTCTCAAGCTGCAATGCTGTTCAAGCCAAA 3652
 Db 3571 GCAATCAAGGCTTCAATTTCTAGGCGGAGCGAGTGAATTCGAGACCAATCTAGGA 3630
 QY 3653 CTGGCTACCTGAGCCCTGTTATGGGACGAGCGGCTTTCTAAAACAAAACAGACAA 3712
 Db 3631 CTGGCTACCTGAGCCCTGTTATGGGACGAGCGGCTTTCTAAAACAAAACAGACAA 3690
 QY 3713 CAACAGCAATTTTAACTGGAAGTGGCTTCAAAATATTAATTAATGAGGCTGATTCAT 3772
 Db 3691 CAACAGCAATTTTAACTGGAAGTGGCTTCAAAATATTAATTAATGAGGCTGATTCAT 3750
 QY 3773 CATCAACCTGGAAGTGGCTTCAAAATATTAATTAATGAGGCTGATTCAT 3832
 Db 3751 GGTGAATC-----CGGCTCAATGCAAGCCCAAGAGAGTGAAGAAAGTTTCTCTCA 3804
 QY 3833 GAGCGGTGATGATTTTGGAAAGAGAGCGCGGAGCTTCAAACTGATTTGACAA 3892
 Db 3805 GAGCGGTGATGATTTTGGAAAGAGAGCGCGGAGCTTCAAACTGATTTGACAA 3864
 QY 3893 TGTATGATTTAAGAGAGAGAAATTAAGCACTAATCCCTGGGCAACCGAAAGATT 3952
 Db 3865 GGTATGATTTAAGAGAGAGAAATTAAGCACTAATCCCTGGGCAACCGAAAGATT 3924
 QY 3953 TGGGACCGTGGAGTCAATTTCCAGAGCAGACAGACCTTGGCAACCGAGATGGA 4012
 Db 3925 TGGTCTGTATCTACCAACCTCCAGAGAGCAACAGACAGCTACCGCAATGTCAA 3984

QY 4013 TGCTATGGAGCATTAACCTGGGATGAGTGGGCAAGATAGAGAGCTGACGGGTC 4072
 Db 3985 CACACAAAGGCTTCTTCAAGCATGCTGCGAGACAGAGATGATCTTCAAGGGCC 4044
 QY 4073 CATTTGGGCAAAATTTCTTCAACAGATGAGCACTTTCACCCCTCTCTTATGAGCG 4132
 Db 4045 CATCTGGGCAAAATTTCTTCAACAGATGAGCACTTTCACCCCTCTCTTATGAGCG 4104
 QY 4133 CTTTGGACTCAAGAACCCGCTCTCTCAAGATCTCATCAAAAACAAGCTGTTCTGCGAA 4192
 Db 4105 ATTGCACTTAAACACCTCTCTCAAGATCTCATCAAAAACAAGCTGTTCTGCGAA 4164
 QY 4193 TCTCGGCGGAGATTTTCAAGCTCAAGATTTGCTTCAATCAACCAATATCTCCAGG 4252
 Db 4165 TCTTGGACCACTTCAAGTGGGCAAAAGTTGCTTCTTCAATCAACAGATCTCAAGG 4224
 QY 4253 ACA-AGTGAATGGAATTTGAATGGAGCTGCAAGAAAGAAACAGCAAGCTGGAATC 4311
 Db 4225 ACAAGGTCAAGCTGGAATGGAATGGAGCTGCAAGAAAGAAACAGCAAGCTGGAATC 4284
 QY 4312 CCGAAGTCAAGTCAATCAATCAATTTATGCAAAATCTGCAAGCTTATTTTATCTGACA 4371
 Db 4285 CCGAATTCAGTCACTTCACTCAACAGCAAGCTGTTAATCGTGAATTTACCTGGAAT 4344
 QY 4372 ACAATGAGCTTTATCTGAGGCTCGCCGATTTGGACCCGTTACCTTACCCGCTCTG 4431
 Db 4345 CTATAGCGGTATTAAGAGCTCGCCGATTTGGACCCGATTTACCTTACCTTACCTG 4404
 QY 4432 AATTAGCTTATTAATTAATTAATCGGTTGATTCGTTCAAGTGAATTTGCTCTGCTC 4491
 Db 4405 AATTGCTTATTAATTAATTAATCGGTTGATTCGTTCAAGTGAATTTGCTCTGCTC 4464
 QY 4492 TTCTTATCTTATGCTTACAGTGTATTAAGCTTACATTAATCTGTTGCTGCTG 4551
 Db 4465 TTCTTATCTTATGCTTACAGTGTATTAAGCTTACATTAATCTGTTGCTGCTG 4524
 QY 4552 CGATTAAGACTTACGATCAAGCTGTTACCTGTTATGAGTGAATGAGTGGCCACTGCTG 4611
 Db 4525 ACTCAAGGA-----ACCCCTATGATGATGAGTGGCCACTGCTGCTG 4568
 QY 4612 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4671
 Db 4569 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4628
 QY 4672 CTGCGGCTTCAAGGAGCGAGCGGCGGAGAGAGAGTGGGCA 4718
 Db 4629 CGGCGGCTTCAAGTGAAGCGAGCGGCGGAGAGAGAGTGGGCA 4675

RESULT 10
 ABA02989
 ID ABA02989 standard; DNA; 4675 BP.
 XX
 AC ABA02989;
 XX
 DT 19-FEB-2002 (first entry)
 XX
 DE Adeno-associated virus 2 genome DNA sequence SEQ ID NO 1.
 XX
 DE Cytostatic; virocid; apoptosis inducer; p53; Saos-2 cell; infection;
 KW cancer; virus; HPV16; HPV18; adeno-associated virus 2; AAV-2; genome; BS.
 OS Adeno associated virus 2.
 PN W0200180840-A2.
 XX
 PD 01-NOV-2001.
 XX
 XX 20-APR-2001; 2001WO-GB01795.
 XX
 PR 20-APR-2000; 2000GB-0009887.
 XX

PA (BTG-) BTG INT LTD.
 XX Raj K. Beard PM.
 XX WPI; 2002-041365/05.
 XX
 PT Single stranded and/or looped DNA for treating mutant p53 associated
 PT cancer or infection that inhibit cellular p53 having portion with an
 PT base, internally located with respect to any 3' and 5' ends of the DNA
 XX
 PS Disclosure; Page 36-38; 51pp; English.
 XX
 CC The invention relates to single stranded and/or looped DNA having a
 CC portion with at least one base, internally located with respect to any 3'
 CC and 5' ends of the DNA, that is unbasepaired with another base in a form
 CC that is capable of being internalised within a target cell, for use in
 CC therapy, with cytostatic and virucide activity. The DNA acts as an
 CC apoptosis inducer in cells that lack p53 functionality, useful for
 CC killing a cell, preferably a dividing cell where the cell is other than a
 CC Saos-2 cell and the DNA is not configured to express the peptide or
 CC protein that selectively kills the cell. The DNA is preferably in the
 CC form of an AAV or associated with AAV protein which has been treated such
 CC that the DNA is no longer capable of replication or expression in cells
 CC and is associated with or contained within a vehicle which is associated
 CC with one or more viral fibers which facilitate internalisation of the DNA
 CC into a target cell. The DNA is also useful for manufacturing a medicament
 CC for treating an individual suffering from a mutant p53 associated cancer
 CC or an infection that inhibits cellular p53. The method targets cancer
 CC cells or cells infected with p53 inhibiting viruses, such as HPV16 or
 CC HPV18 and only cells that lack p53 activity are killed and no damage to
 CC cellular DNA is involved. The present sequence is that of the
 CC adeno-associated virus (AAV) 2 genomic DNA sequence.
 CC
 XX
 SQ Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;
 Query March 64.8%; Score 3055.8; DB 24; Length 4675;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

463 TTGAGCAGGACACCTCGACCGTGGCCGAGAAAGCTGACGGCAGCTTTCTGACGGAATGCG 522
 537 GCCCGGTAGTAAGGCCCCGGGAGGCCCTCTTTTGTAGTTCGAGAAAGGCGAATCCT 596
 523 GCCGTGTAGTAAGGCCCCGGGAGGCCCTTTTGTGCAATTTGAGAAAGGAGAGACT 582
 597 ACTTCCACCTCCATTTCTGTGTGAGAGACCGAGGGGTCAATTCATGATGCTGGCGCT 656
 583 ACTTCCACATGACGCTGCTGTGTAACCAACCGGGGTGAAATCATGTTTGGAGCTT 642
 657 TCCTGATCAGATTAGGAGCAACGCTGTGCAACATCTACCGCGGATGAGCCGACCC 716
 643 TCCTGATCAGATTAGGAGCAACGCTGTGCAACATCTACCGCGGATGAGCCGACCT 702
 717 TGCCTCAATGCTTGGGGGTGACCAACGCGTAAATGGCCCGGAGGGGAGAAAGGTG 776
 703 TGCCTCAATGCTTGGGGGTGACCAACGCGTAAATGGCCCGGAGGGGAGAAAGGTG 762
 777 TGGACGAGTCTACATCCCACTACCTCTGCGCCAAAGCTCAGCCGAGCTCAGTGG 836
 763 TGAATGAGTCTACATCCCACTACCTCTGCGCCAAAGCTCAGCCGAGCTCAGTGG 822
 837 CGTGAATCAATGAGAGAGATATATAAGCGCTGTTTAACTTGGCCGAGCCGAAACGCG 896
 823 CGTGAATCAATGAGAGAGATATATAAGCGCTGTTTAACTTGAATCTCAGCGAGCTTAAACGCT 882
 897 TCGTGGCGGACGACCTGACCCAGCTGACGACGACGACGACGACGACGACGACGACGAC 956
 883 TGTGTGGGAGGACGACCTGACCCAGCTGACGACGACGACGACGACGACGACGACGACGAC 942
 957 ACCCCAAATTTGACGCGCGCTGTATCCGCTCAAAACCTCCGCGCTACATGAGAGTGG 1016
 943 ATCCCAATTTGATGAGCGCGCTGTATCCGCTCAAAACCTCCGCGCTACATGAGAGTGG 1002
 1017 TCGGCTGCTGTGTGACACCGGGGACATCACTCCGAGAAAGAGTGAATCAGAGAGACCAAG 1076
 1003 TCGGCTGCTGTGTGACACCGGGGACATCACTCCGAGAAAGAGTGAATCAGAGAGACCAAG 1062
 1077 CTTGTGATCTCTCTTCAACGCGGCTTCAACCTCGGGTCCGAGTCAAGGCGCTCTGG 1136
 1063 CTTGTGATCTCTCTTCAACGCGGCTTCAACCTCGGGTCCGAGTCAAGGCGCTCTGG 1122
 1137 ACAATGCGGCAAGATCATGAGCGCTGACCAATTCGCGCCGACTACCTGTGAGGCCCG 1196
 1123 ACAATGCGGCAAGATCATGAGCGCTGACCAATTCGCGCCGACTACCTGTGAGGCCCG 1182
 1197 CTCGCGCGGCAAGATCATGAGCGCTGACCAATTCGCGCCGACTACCTGTGAGGCCCG 1256
 1183 AGCCCGTGAAGACATTTCCAGCAATCGAATTTATTAATTTTGAACCTTAACGGGTACG 1242
 1257 AACCTGCTACGCGCGGCTCGCTCTTCTGTGAGTGGGCGGCAAAAGGTTGGGAGCGCA 1316
 1243 ATCCCAATTTGAGCGCTTCTCTTCTGTGAGTGGGCGGCAAAAGGTTGGGAGCGCA 1302
 1317 ACAATCTGCTGTGTTGGGCGGCAACCAAGGAGCAACCAATCTGCGGAAAGCAATCG 1376
 1303 ACAATCTGCTGTGTTGGGCGGCAACCAAGGAGCAACCAATCTGCGGAAAGCAATCG 1362
 1377 CCGACGCGGCTGCTTCTACGCTGCTGCACTGACCAATGAGAACTTCTCTCAATG 1436
 1363 CCGACGCGGCTGCTTCTACGCTGCTGCACTGACCAATGAGAACTTCTCTCAATG 1422
 1437 ATTGCGTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1496
 1423 ACTGTGTCAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1482
 1497 AGTCCGCAAGGCAATTTCTGCGGAGCAAGAGTGGCGGTGACCAAAATGCAAGTCTG 1556
 1483 AGTCCGCAAGGCAATTTCTGCGGAGCAAGAGTGGCGGTGACCAAAATGCAAGTCTG 1542
 1557 CCGCGCAAGTCAAGCCCAACCCCGTGTATCTCACTTCAACCAACCAATCTGTGCGCGTGA 1616

1543 CGGCCAGATAGACCCGACTCCCGTATCTCACTTCAACCAACCAATATGCGCCGTGA 1602
QY 1617 TTGACGGGAACACACACCTTTGAGACACAGAGCCCTTGGAGACCGGATTTTCAAT 1676
Db 1603 TTGACGGGAACCTCAAGACCTTGAACACAGAGCCCTTGGAGACCGGATTTTCAAT 1662
QY 1677 TTGAACTCACCCCGCTTGGAGACATGCTTTGGCAAGTGACAAAGCAGAGATCAAG 1736
Db 1663 TTGAACTCACCCCGCTTGGAGACATGCTTTGGCAAGTGACAAAGCAGAGATCAAG 1722
QY 1737 AGTTCTTCCGCTGGGCGCAGATGACGTGACCGAGGTGGCGATGATGCTTCAAGAA 1796
Db 1723 ACTTTTCCGCTGGGCGCAGATGACGTGATGAGGTGAGCATGATTTCTAGTCAAAA 1782
QY 1797 AGGTGAGGACCAAAAGACCCGCCCCGATGACCGGATTAAGCCGACCCCAACCGGG 1856
Db 1783 AGGTGAGGACCAAAAGACCCGCCCCGATGACCGGATTAAGGTAGAGACCCCAACCGGG 1842
QY 1857 CCTGCCCTCAGTCGCGGATTCATCGACGTGACAGCGGAGAGGCTCCGGTGAACCTTTG 1916
Db 1843 TGGCGAGTCAGTTGGCGAGCCATCGACGTGACAGCGGAG--AGCTTGCATCAACTAG 1899
QY 1917 CCGACAGGTAACAAAATGTTCTGTCACGCGGCGATGCTTCAAGATGCTTTCCCT 1976
Db 1900 CAGACAGGTACAAAATGTTCTGTCACGTTGGGCGATGATGATGCTTTCCCT 1959
QY 1977 GCAGACATGCGGAGAAATGAAATCAGAAATTTCAACTTTGCTTCAAGCAGGACAGAG 2036
Db 1960 GCAGACATGCGGAGAAATGAAATCAGAAATTTCAACTTTGCTTCAAGCAGGAGAAAG 2019
QY 2037 ACTGTTCAAGAGTCTTCCCGGCGGTGACAGATCTCAACCGGT---CGTCAGAAAGAGA 2093
Db 2020 ACTGTTCAAGAGTCTTCC---CGTCAGAAATCTCAACCGCTTTCTGTCGTAAGAAAG 2076
QY 2094 CGTATCGAAACTCTGTGCCATTCATCTGCTGTGGGCGGAGCTCCGAGATGCTTGTCT 2153
Db 2077 CGTATCGAAACTCTGTGCCATTCATCTGCTGTGGGAGAGGTGCCAG---CGCTTGA 2133
QY 2154 CGGCTTGGCATCTGTGTCACAGTGAACCTGATGATGCTGTCTTCTGAGCAATTAATGACTT 2213
Db 2134 CTGCTGCGCATCTGTGTCACAGTGAATTTGATGATGATCTTCTTGAACAAATTAATGATTT 2193
QY 2214 AAACAGATATGCTGCGCATGCTTATCTTCCAGATTTGCTCGAGGACAACTCTCTGAG 2273
Db 2194 AAATCGAGTATGCTGCGCATGCTTATCTTCCAGATTTGCTCGAGGACAACTCTCTGAA 2253
QY 2274 GGCATTCGCGAGTGTGGGACTTGAAACTGAGCCCGGAGCCCAAGCCCAACCAAGCA 2333
Db 2254 GGAATTAAGACATGTGGGAGGCTCAAACTGGCCCAACCAAGCCCGGAGCCG 2313
QY 2334 AAGCAGAGCAGCGCGGGGTCTGTGTCTTCTGTGCTACAGTACTCTCGACCTTCAAC 2393
Db 2314 CATTAAGACAGCAGCGGGGTCTGTGTCTTCTGTGCTACAGTACTCTCGACCTTCAAC 2373
QY 2394 GGCATCGACAAAGGGGAGCCGTCACAGCGGCGGAGCGGAGCCCTCGAGCAG---ACAA 2452
Db 2374 GGCATCGACAAAGGGGAGCCGTCACAGCGGCGGAGCGGAGCCCTCGAGCAGTACAA 2433
QY 2453 GGCCTACGACACAGCAGCTCAAAAGCGGCTGACAACTCGATATTAACACGCGCA 2512
Db 2434 AGCCTACGACCGGAGCTGACAGCGGAGCAACCGGTACTCAAGTACAAACGCGCA 2493
QY 2513 CCGCGAGTTTCAAGAGGCTTGAAGAAATAGCTTTTGGGGGCAACTCGGGGAGC 2572
Db 2494 CCGGAGTTTCAAGAGGCTTGAAGAAATAGCTTTTGGGGGCAACTCGGAGAGC 2553
QY 2573 AGCTTTCAGAGGCAAGAGGCTTGAAGCTTCTGAGCTGCTGATGAGAGAGCTTAA 2632
Db 2554 AGCTTTCAGAGGCAAGAGGCTTGAAGCTTCTGAGCTGCTGATGAGAGCTTAA 2613
QY 2633 GACGCTCTCTGAGAAAGAGCTCGGTGAGCAGTGCACAAAGCAGACCTCTCTC 2692
Db 2614 GACGCTCTCTGAGAAAGAGCTCGGTGAGCAGCTCTCTGAGCAGACCTCTCTC 2673

2693 GGGCATCGGCAAGACAGGCGCAGAGCCGCTAAAAAGAGACTCAATTTTGTGACAGCTG 2752
QY 2674 GGGATCGGAAAGGCGGGCCAGAGCTGCAAGAAATTAATTTTGTGACAGCTG 2733
Db 2753 CGACTCAGATCAGTCCCGATTCACAACTCTCGGAGAACTCCAGCAACCCCGCTGC 2812
QY 2734 AGACGAGACTCAGTACCTGACCCCGAGCTCTCGGAGAGCCAGCAGAGCCCGCTG 2793
Db 2813 TGTGGGACTTACTCAATAGGCTTCAAGCGGTGGCGCAACAAATGAGCAATTAAGAG 2872
QY 2794 TGTGGGAACTTAATCAGTATGCTTCAAGCAGTGGCGCAACAAATTAAGAGAG 2853
Db 2873 CGCGGAGGAGTGGGTAATGCTTCAAGAAATTTGGCTTGCATTTCCAGTGGCGGCA 2932
QY 2854 CGCGGAGGAGTGGGTAATGCTTCCGAAATTTGGCTTGCATTTCCAGTGGGCGA 2913
Db 2933 CAGAGTCATCACACAGCAGACCCGACCTTGGGCTTGCACCTTCAATTAACCACTCTA 2992
QY 2914 CAGAGTCATCACACAGCAGACCCGACCTTGGGCTTGCACCTTCAATTAACCACTCTA 2973
Db 2993 CAAAGCAATCTCAGTGTCTTCAACGGGGGCGCAGCAACCAACCTACTTGGCTACG 3052
QY 2974 CAAACAAATTTCCAGGCAATCA---GAGCTCTGAACGAACTACTCTTGGCTACAG 3030
Db 3053 CACCCCTGGGGGTAATTTGATTTCAAGATTTCCATGCGCACTTTTCAACGCTGACG 3112
QY 3031 CACCCCTGGGGGTAATTTGATTTCAAGATTTCCATGCGCACTTTTCAACGCTGACG 3090
Db 3113 GAGGCACTCATCAACCAATTTGGGATTTCCGGCCCAAGAGACTCAACTTCAACTT 3172
QY 3091 GCAAGACTCATCAACCAATTTGGGATTTCCGGGATTTCCGACCCAGAGACTCAACTT 3150
Db 3173 CAACATCAAGTCAGAGAGTCAACAGATATGAGCTTCAACCACTTGTATTAATCT 3232
QY 3151 TAACTTCAAGTCAGAGAGTCAACAGATATGAGCTTCAACCACTTGTATTAATCT 3210
Db 3233 TACAGCAGGTTTCAAGTCTTCTCGGACCTGAGTACAGCTTCTGTCGCTGCGCTC 3292
QY 3211 TACAGCAGGTTTCAAGTCTTCTCGGACCTGAGTACAGCTTCTGTCGCTGCGCTC 3270
Db 3293 TGGCAGCAGGCGCTGCTCCCTCCGTTCCCGGCGGAGCTGTTCAATGATTCGCAATACG 3352
QY 3271 GGGGATCAAGAGATGCTCCCGCGTTCCAGAGAGCTTCAATGATTCGCAATACG 3330
Db 3353 CTACTGACGCTCAACATGAGCAAGCCGCTGGAGCTTCAATGATTCGCAATACG 3412
QY 3331 ATACTCAGCTGAGCAAGGAGTCAAGGAGTCAAGGAGCTTCAATGATTCGCAATACG 3390
Db 3413 ATATTTCCCTTCTCAGATGCTGAGAAAGCGGCAACCTTACTTCACTTCACTTGA 3472
QY 3391 GTACTTCTTCTCAGATGCTGAGAAAGCTTACTTCACTTCACTTGA 3450
Db 3473 GGAAGTCCCTTCTCAGAGAGCTACGCGCAGCAGAGCTTGAACCGGCTGATGATCC 3532
QY 3451 GGAAGTCCCTTCTCAGAGAGCTACGCGCAGCAGAGCTTGAACCGGCTGATGATCC 3510
Db 3533 TCTCATCGACCAATACCTGATTAAGCAAGAACTTCACTTCACTTGA 3592
QY 3511 TCTCATCGACCAATACCTGATTAAGCAAGAACTTCACTTCACTTGA 3570
Db 3593 AAAAAGAGCTTGTGTTAGCGGTGCTCAGTGGCAATGCTTCACTTGA 3652
QY 3571 GCGATCAGGCTTCAATTTTCTCAGGCGGAGCGAGTCACTTGGAGCAGCTTGA 3630
Db 3653 CTGGCTACTGAGACCTGTTATCGGAGAGCGGCTTTTAAACAAAGACAGCAACAA 3712
QY 3631 CTGGCTTCTGAGACCTGTTATCGGAGAGAGATTAAGAACTTGGGATTAACA 3690
Db 3713 CAACAGCAATTTTACTGAGCTTGTCTTCAAAATTAATCTCAATGAGCTTGAATCAT 3772
QY 3691 CAACAGTGAATCTGCTGAGCTGAGCTTCAAGTACCACTCAATGAGGAGACTTCT 3750

```

QY 3773 CATCAACCTGGACACTGCTATGCTCACAACAAGACGAGACAGTCTTCCCAT 3832
DB 3751 GGTGAATC-----CGGCAATGGCAAGCCACAGAGAGTGAAGAAATTTTCTCA 3804
QY 3833 GAGCGGTGTCATGATTTTGGAAAAGAGCGCGAGCTTCAAACTGCAATGACAA 3892
DB 3805 GAGCGGGTCTCTCATCTTTGGGAAGCAAGCTCAGAAACAAATGTGAACATTTGAAA 3864
QY 3893 TGTGATATTAACAGACGAAGGAAATTAAGCACTAACCTGTGGCCACCGAAGATT 3952
DB 3865 GGTATATTAACAGACGAAGGAAATCGAACAACCAATCCGTGGCTACGGAGCAGTA 3924
QY 3953 TGGGACCGTGGACGTCAATTTCCAGACAGACGACGACCTTCCGACCGGAGATGCA 4012
DB 3925 TGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACGCTACCGGAGATGCA 3984
QY 4013 TGTATGGAGACATTACCTGGCATGTGTGGCAAGATAGACGCTGACCTGACGGTCC 4072
DB 3985 CACACAAGCGCTTCTCAGGCAATGTGTGGCAAGATGTATCCTTCAGGGGCC 4044
QY 4073 CATTTGGGCAAAATTTCTCAACAGATGACACTTTACCCCTCTCTTAATGGGCGG 4132
DB 4045 CATCTGGGCAAAAGTTCCACAGACGACGACATTTTACCCCTCTCTCATGGGTGG 4104
QY 4133 CTTTGACTCAAGAACCGGCTCTCAGATCCTCATCAAAAACAGCGCTGTCTCGGAA 4192
DB 4105 ATTGCACTTAACACCTCTCTCAGATCTCTCATAGAACACCCCGGACTGTGAA 4164
QY 4193 TCTCGCGGAGTTTCACTCAAAAGTTTCTTCAATTCATCAACCAATATCTCACAGG 4252
DB 4165 TCTTCCACACACTTCAGTGGGCAAAAGTTTCTTCTTCAACAGATCTCACAGGG 4224
QY 4253 ACA-AGGAGGTGGAAATTAAGGAGCTGCGAAGAAAGAAAGAGGCGTGGATC 4311
DB 4225 ACAGCGTACAGGTGAATGAGTGGAGCTGCGAAGAAAGAAAGAGGCGTGGATC 4284
QY 4312 CCGAAGTGCAGTACACATCCATTAATGCAAAATCTGCCAAGTGTATTTACTGTGACA 4371
DB 4285 CCGAATTTAGTACACTTCCACATCAACAAAGTGTATTAATCGTGACTTACCGTGATA 4344
QY 4372 ACAATGACTTATATGAGCTTCCGCGCATTTGGCACCCGTTACCTTACCCGCTGT 4431
DB 4345 CTAAATGCGTGTATGAGAGCTTCCGCGCATTTGGCACAGATCTGATATCTGT 4404
QY 4432 AATTAGCTGTAATCAATTAACCGGCTGATTCGTTAGTGAACCTTGTCTCTCTCC 4491
DB 4405 AATTGCTGTATCAATTAACCGGCTGATTCGTTAGTGAACCTTGTCTCTCTCC 4464
QY 4492 TTCTTATCTTATCGGTTACCATGTTATAGCTTACATTAACCTGTGGTGGCTTCC 4551
DB 4465 TTCTTATCTTATCGGTTACCATGTTATAGCTTACATTAACCTGTGGTGGCTTCC 4524
QY 4552 CGATTAAGACTTACGTCATCGGTTACCCCTAGTATGAGTGGCCACTCTCTCTG 4611
DB 4525 ACTACAAAGG-----ACCCCTAGTATGAGTGGCCACTCTCTCTCTG 4568
QY 4612 CGGCGTGTGTGCTCGGTTGGGGCTGGGGAACCAAGGTCCGACGAGAGCTGTGCT 4671
DB 4569 CGGCGTGTGTGCTCGGTTGGGGCTGGGGAACCAAGGTCCGACGAGAGCTGTGCT 4628
QY 4672 CTGCGGCGCCCAACGAGCGGACGCGGACGAGAGGAGAGTGGGCA 4718
DB 4629 CGGCGGCGCTCAGTGAAGGAGCGGCGGAGAGGAGAGTGGGCA 4675

```

RESULT 11
AAH26326
ID AAH26326 standard; DNA; 7557 BP.

XX AAH26326;
XX
XX 02-OCT-2001 (first entry)
XX

```

DE Adeno-associated virus packaging plasmid pflxOAAV.
XX
XX AAV; vector; pflxOAAV; packaging cell line; gene therapy; ds.
KW
XX Adeno associated virus.
OS
XX MO200155361-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 26-JAN-2001; 2001MO-US02709.
PF
XX
XX 26-JAN-2000; 2000US-0178536.
PR
XX
XX (CHIR ) CHIRON CORP.
PA
XX
XX Hardy SF;
PT
XX WPI; 2001-483239/52.
DR
XX
XX Producing recombinant adeno-associated virus (rAAV) vector, by stably
PT infecting eukaryotic host cell with rAAV vector, helper adenovirus or
PT herpes virus, recombinant adenovirus or herpes vectors -
PS
XX Example 1; Page 54-57; 63pp; English.
XX
XX The present sequence is that of adeno-associated virus (AAV)
XX packaging plasmid pflxOAAV2. The following AAV packaging plasmids
XX were prepared: pKStrepap, a non-replicating control plasmid
XX containing rep and cap genes but no inverted terminal repeats
XX (ITRs); pAV2 (see AAH26325), containing a wild-type AAV genome; and
XX pflxOAAV, a version of pAV2 with 2 loxp sites inserted such they
XX flank the rep and cap genes and thus separate the ITRs from these
XX genes. Each was combined with pcWV GFP (see AAH26324), a plasmid
XX with a green fluorescent protein expressing recombinant AAV (rAAV)
XX genome, and used to transfect 293 or Cre8 cells. After 6 hr, the
XX medium was changed and wild-type adenovirus type 5 was added.
XX Virus particles were harvested 3 days later. Plasmid pAV2
XX produced mostly AAV and a low yield of GFP vector with no effect
XX of Cre recombinase. pflxOAAV packaged as much GFP vector as the
XX control plasmid and while the amount of GFP vector was not
XX affected by Cre recombinase, the amount of pflxOAAV in the product
XX was reduced to 1/10 of the 293 value by the action of Cre
XX recombinase. In 293 cells it was evident that the GFP vector did
XX not compete with wild-type AAV for replication or packaging, while
XX the GFP vector was able to compete against pflxOAAV. Methods
XX and compositions are provided for producing rAAV vector particles
XX by: (a) introducing into a host cell (1) pflxOAAV, (1i) a
XX recombinant viral vector encoding plasmid, and (1ii) a plasmid
XX encoding herpesvirus, cytomegalovirus or adenoviral functions, or
XX a herpesvirus, cytomegalovirus or adenovirus itself, to produce
XX flox AAV particles and rAAV particles; and (b) introducing into a
XX second host cell (1) the rAAV particles or (a), (1i) a vector
XX that directs expression of Cre, and (1i) a vector which directs
XX expression of herpes virus, cytomegalovirus or adenovirus helper
XX functions, such that rAAV vector particles are produced. The
XX vectors are useful for in vivo or in vitro gene therapy and also
XX for in vitro recombinant protein production.
XX
XX Sequence 7557 BP; 1970 A; 1940 C; 1953 G; 1694 T; 0 other;
SQ

```

Query Match 63.4%; Score 2993.4; DB 22; Length 7557;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 3633; Conservative 0; Mismatches 841; Indels 45; Gaps 9;

```

QY 18 GCGGCTGCTGCTGCTGCTGCTGCTGCGGACCAAGTCTCCGACAGCGGACGCTCTGC 77
DB 4 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCGGACCAAGTCTCCGACAGCGGACGCTCTGC 63
QY 78 TCTGCGGCGCCCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 137
DB 64 TCGCGCGGCTCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 123

```

OY	128	GGGTAT-----CGGAAAGCGCTCCACGCTGCCGCTC	172
Db	124	TGGACTTGTACATTACTTGTAATGTATATGATACGAAGTTATCCGAGGGGTGG	183
OY	173	AGCGGTGACGTAAATTACGTCAATAGG---GAGTGGTCTGTATTAGCTGCACGTGAGT	229
Db	164	AGTCGTGAGTGAATTACGTCAATAGGTTTAAAGAGAGGTACGTGAGT	243
OY	230	GCTTTTGGACATTTTGGACACCACTGGCCATTTTAAAGGTATATATGAGCCGAGTGAAGC	289
Db	244	G-TTTTGGACATTTTGGACACCACTGTGTACGCTGGGTATTTTAAAGCCGAGTGAAGCA	302
OY	290	AGCAGATTCCTCATTTTG-ACCGCCAAATTTTGAACGAGACGACCTACCGGCTTCTTA	348
Db	303	CGACAGGTTCTCATTTTGAAGGGAGGTTTAAAGCCGACGCCCATATCCGAGGTATTTTA	362
OY	349	CGAGATCGTATCAAGGTGCGGAGGCACTTGAACGACCTGCGGGGACTTCTGACTC	408
Db	363	CGAGATTGTGATTTAAGTCCCGACGACCTTGAACGACCTTGCCCGGATTTCTGACAG	422
OY	409	GTTTGTAGCTGGGTGGCCGAGAAAGATGGAAGCTGCCCGGATTTCTGACATGATCT	468
Db	423	CTTTGTGAATCGGTGGCCGAGAAAGATGGAAGTGGAGTTGCCGACATTTCTGACATGATCT	482
OY	469	GAATCTGATTTGACGACGACCCCTTACCGTGGCCGAGAAAGCTGACGCGGACTTCTGTGT	528
Db	483	GAATCTGATTTGACGACGACCCCTTACCGTGGCCGAGAAAGCTGACGCGGACTTCTGTGAC	542
OY	529	CCAAATGCGCGCGGTGAGTAAAGGCCCGGAGGCGCTCTTCTTGTTCAGTTGAGAGAGG	588
Db	543	GGAAATGCGCGCGGTGAGTAAAGGCCCGGAGGCGCTCTTCTTGTTCAGTTGAGAGAGG	602
OY	569	CGAGTCTTACTTTCACCTTCATATTTCTGGTGGAGACCAAGGGGGTCAATTCATGATGTCT	648
Db	603	AGAGAGTACTTCCACATGCAAGTGTCTGGTGAACCAACGGGGTGAATTCATGATGTCTT	662
OY	649	GGGCGGCTTCTTGAGTCAATTAAGGACAAAGCTGGTGAACCAATCTACCGGGATGA	708
Db	663	GGGAGTTTCTCGATCGATTCAGATTCGGAAAMACGTGATCAGAAATTTTACCGGGGATGA	722
OY	709	GCCGACCTTGGCCCACTGAGTTTGGCGGTGAACAAAGACGGTAAATGGCGCGAGGGGGAA	768
Db	723	GCCGACCTTGGCCCACTGAGTTTGGCGGTGAACAAAGACGGTAAATGGCGCGAGGGGGAA	782
OY	769	CAAGGTGTGACAGAGTGTACATCCCACTACCTCTGCGCCCAAGACTCAGCCCCGAGCT	828
Db	783	CAAGGTGTGACAGAGTGTACATCCCACTACCTCTGCGCCCAAGACTCAGCCCCGAGCT	842
OY	829	GCAGTGGGCTGTGACTAACTGAGAGATATATTAAGCGCTGTTTGAACCTTGGCCGAGCG	888
Db	843	CCAGTGGGCTGTGACTAACTGAGAGATATATTAAGCGCTGTTTGAATCTCACGAGACG	902
OY	889	CAAAAGCGCTCGTGGCGAGCACCTGACCAAGCTGACGACCCGAGGCGAGAACAGGA	948
Db	903	TAAACGTTGGTGGCGAGCATCTGACGACCTGTGCGAGACGACGAGACGAAPAAAGA	962
OY	949	GAATCTGAACCCCAATTTGACGCGCTGTCAATCCGATCAAAAACCTCGCGCTGATAT	1008
Db	963	GAATCAAGATCCCAATTTGATGATGCGCGCTGTCAATCAAAAACCTTCAAGCAAGTATCAT	1022
OY	1009	GGAAGTGGTGGGTGGCTGTGAGCAAGGGGATTTACTCTGAGAAAGAGTGTATCAAGA	1066
Db	1023	GGAAGTGGTGGGTGGCTGTGAGCAAGGGGATTTACTCTGAGAAAGAGTGTATCAAGA	1082
OY	1069	GGAACGAGGCTGTGATCTTCCCTTCAAGCGCGCTTCAAACTCGGGGTCCCAATGAAGC	1122
Db	1083	GGAACGAGGCTGTGATCTTCCCTTCAATGCGGCTTCAAACTCGGGGTCCCAATGAAGC	1142
OY	1129	CGCTCTGACAAATGCGGCAAGATCATGGCTGACCAATTCGCGCCGACTACCTGT	1188
Db	1143	TGCTTGGACAAATGCGGGAAGAATTATGAGCTGTAAACCCGCCGACTACCTGT	1202
OY	1189	AGGCCCGGCTCGCCCGGAGCATTTAAACCAACCGCATCTACCGCATCTGGAGCTGA	1248

Db	1203	GGGCGAGAGCCCGTGGAGGAGCATTTCCAGCATGAGATTATTAATTTTGGAACTAAA	1267
Qy	1249	CGGCTACGAAACCTGGCTTAAGCGCGCTCGTCTTTCTCGGCTGGGCCAGAAAAGGTTCGG	1300
Db	1263	CGGGATACGATCCCAATATGCGGCTTCGCTCTTTCTGGGATGGCCAGAAAAGTTTCGG	1322
Qy	1309	GAAAGCGCAACACCATCTGGCTGTTTGGCCCGGCCAACACGCGCAAGACCAACATCGCGGA	1366
Db	1323	CAAGAGGAAACCATCTGGCTGTTTGGGCTGCACTACCGGGAGAGCAACATCGCGGA	1387
Qy	1369	AGCCATTCGCCCAACGCCGTGCTTTCTAAGGCTGCTCACTGAGCAATGAACTTTTC	1422
Db	1383	GGCCATAGCCCAACATGTGGCTCTTTACGGGGCTTAACTGGACCAATGAAACCTTTTC	1443
Qy	1429	CTTCATGATTTGCTGTCGACAAGATGGTATCTGTGGGAGGAGGCAAGATGACGGCCAA	1488
Db	1443	CTTCAACGACTGTGTGACAGATGTGATCTGTGGGAGGAGGAGGAAAGATACCGCCCA	1507
Qy	1489	GGTGTGAGAGTCCGCGCAAGGCCATTCTCGCGCGCAGCAAGTGCGCGTGAACAAAAGTG	1544
Db	1503	GGTGTGAGAGTGGCGCAAGGCATTCTCGAGAGAGCAAGGAGCGCGTGAACCAAAATG	1565
Qy	1549	CAAGTGTTCGCCCGCATGACATCCCAACCCCGTGATGCTACCTTCAACCAATGTG	1607
Db	1563	CAAGTGTTCGGCCGCAATGACCCGACTCCGTTGTCATCTTCAACCAACCATGTG	1622
Qy	1609	CGCCGTGATTTGACGGGAAACAGACCAACCTTCGAGACAGACGCGGTTCAGACCGGAT	1666
Db	1623	CGCCGTGATTTGACGGGAACTACAGCACTTTGGAACACAGACGCGGTTCAGACCGGAT	1687
Qy	1669	GTTCAAAATTTGAACCTCACCCGCGCTTGGAGATGACTTTGGCAAGTGACCAACAGGA	1728
Db	1683	GTTCAAAATTTGAACCTCACCCGCGCTGATGATCATGACTTTGGGAAGTCAACCAAGGA	1744
Qy	1729	AGTCAAAAGTCTTCCGCTGGGCGCAGATCACTGACCGAAGTGGCGCATGAATTCTA	1788
Db	1743	AGTCAAAACATTTTTCGCTGGGCGCAAGATCACTGTGTTGAGGTGAGCATGAATTTCTA	1802
Qy	1789	CGTCAAAAAGGCTGAGGCAACAAAAGACCCGCGCATGACCGCGATTAAGAGGAGCC	1844
Db	1803	CGTCAAAAAGGCTGAGGCAAGAAAAGACCCGCGCATGACCGCGATTAAGAGGAGCC	1865
Qy	1849	CAAGCGAGGCTGCCCCCTCAGTTCGCGGATCACTGACGTGACGCGGAGAGAGCTCCGAT	1908
Db	1863	CAAAAGGGTGGCGAGTCAGTTTGGCAACCATTCAGCTGACAGCGGGA--AGCTTCAT	1911
Qy	1909	GGACTTTGCCGACAGGTACCAAAACAAATGTTCTGTGACGCGCGGATGCTTCAGATCT	1966
Db	1920	CAACTACCCACAGAGGTACAAAACAAATGTTCTGTGACGCGGCAATGATGATCT	1973
Qy	1969	GTTTCCCTCGAAGACATGCGAGAGATGAATCAGAAATTCAAATTTGCTTCACGCAAGG	2022
Db	1980	GTTTCCCTCGAAGACATGCGAGAGATGAATCAGAAATTTGCTTCACGCAAGG	2033
Qy	2029	GACGAGACTGTTTCAGATGCTTCCCGGCGGTGACAAATCTCAACCGGT---GTCAG	2085
Db	2040	ACAGAAACACTGTTTAAAGTGCTTTC---GTCGCAAACTCAACCCGTTTCTGTGT	2096
Qy	2086	AAAGGAGCGTATCGAAACTCTGTGCATTCAATCTGCTGGGCGGGCTCCGAGAT	2145
Db	2097	CAAAAAGCGTATCGAAACTGTGCTACATTCATTCATGAGGAAAGGTGCGCAGA---	2155
Qy	2146	TGCTTGCCTGGCGTGCATCTGTGTCAGATGTCGACGTCGACGTCGACGATG	2205
Db	2154	CGCTTGCCTGGCGTGCATCTGTGTCAGATGTCGACGTCGACGTCGACGATG	2213
Qy	2206	AATGACTTAAACAGAGTATGCTGCGCATGTTATCTTCAGATTTGCTCGAGGCAAC	2265
Db	2214	AATGATTTAAATCAGGTATGCTGCGCATGTTATCTTCAGATTTGCTCGAGGCACTC	2273
Qy	2266	TCTCTGAGGCGATTTCGCGATGTTGGGACTTGAACCTGAGCCCCGGAAGCTCAAGCCA	2325

2274 TCTCTGAAGATAAGACAGTGTGTGAAGCTCAAACTGGCCCAACCAACCAAGCCG 2333
2226 ACCAGCAAAAGCAGAGACGCGCGGGGTCTGTGCTTCCCTGTACAAAGTAACCTCGGAC 2385
2234 CAGAGCGGCAATAGACACAGCAGGGGTCTTGTGCTTCTGTGGTACAAAGTACCTCGGAC 2393
2286 CTTTCAACGAGCTGACCAAGGGAGCCGCTCAACGCGGCGAAGCAGCGGCTCGAGC 2445
2294 CTTTCAACGAGCTGACCAAGGGAGCCGCTCAACGAGGACAAAGCCGCGGCTCGAGC 2453
2446 ACAGCAAGGCTTACGACCAAGGCTCAAGCGGGTACCAATCCGTACTCTGGGTAAAC 2505
2454 ACAGCAAGGCTTACGACCAAGGCTGACAGCGGAGCAACCCGTACTCAAGTAAAC 2513
2506 ACAGCAAGGCTTACGAGGCTGACAGGAGTACGCTTTGGGGGCAACCTCG 2555
2514 ACAGCAAGGCTTACGAGGCTGACAGGAGTACGCTTTGGGGGCAACCTCG 2573
2566 GCGCAGCAGTCTTCCAGGCGCAAGACCGGGTCTCGAAGCTCTGCTGTGTGAGGAG 2625
2574 GACGAGCAGTCTTCCAGGCGCAAGAGGGTCTTGAAGCTCTGGGCTGTGTGAGGAG 2633
2626 GCGCTAAGACGGCTCTGTGAAGAAAGCTCGGTAGACAGTCCGCAAGAGCGCACT 2685
2634 CTGTTAAGACGGCTCTCGGAAAGAAAGAGCGGGTACGACCTCTGTGTGAGGAGCT 2693
2686 CCTCTCGGGGATCGGGAAGACAGGCGCAGACCGCTTAAGAAAGACTCAATTTGGTC 2745
2694 CCTCTCGGGAACGGAAAGGCGGCGCAGACCTGTCAAGAAAGAAATTTGATTTGGTC 2753
2746 AGACTGCGCACTCAGAGTCAAGTCCCGATCCCAACACTCTCGAGAAACCTTCAGCAACCC 2805
2754 AGACTGCGCACTCAGAGTCAAGTCCCGATCCCGATCCCGATCCCGATCCCGATCCCG 2813
2806 CCGCTGCTGTGGACCTACTCAATGCTTCAAGCGGTGTGGTCAACCAATGGCAGCAATA 2865
2814 CCTCTGTGTGTGGACCTACTCAATGCTTCAAGCGGTGTGGTCAACCAATGGCAGCAATA 2873
2866 ACAGAGGCGCGGAGTGTGGTATGCTCAGGAAATTTGGCAATTTGGCAATTTCCCATGGC 2925
2874 ACAGAGGCGCGGAGTGTGGTATTTCTCTGGGAAATTTGGCAATTTGGCAATTTCCCATGGC 2933
2926 TGGGCGACAGAGTCAATCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2985
2934 TGGGCGACAGAGTCAATCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2993
2986 ACCTCTACAGCAATCTCCAGTGTCTTACGCGGGGCGAGCAAGCAACCACTACTTTCG 3045
2994 ACCTCTACAGCAATCTCCAGTGTCTTACGCGGGGCGAGCAAGCAACCACTACTTTCG 3050
3046 GCTACAGACAGCAGCAGTGTATTTGATTTCAACAGATTCACATGGCACTTTTCAACAG 3105
3051 GCTACAGACAGCAGCAGTGTATTTGATTTCAACAGATTCACATGGCACTTTTCAACAG 3110
3106 GTGACTGCGACAGCTCAATCAACAAATTTGGGATTTCCGCGCCAGAGACTCAACTTCA 3165
3111 GTGACTGCGACAGCTCAATCAACAAATTTGGGATTTCCGCGCCAGAGACTCAACTTCA 3170
3166 AACCTTCAACATCCAGTCAAGAGTCAAGAGATGATGTGGCTCAACCACTTTCGCTA 3225
3171 AGCTCTTAAATTCAGTCAAGAGTCAAGAGATGATGTGGCTCAACCACTTTCGCTA 3230
3226 ATAACTTACAGACAGGCTTCAAGTCTTCTGGAAGTCAAGGATTCAGAGTTCG 3285
3231 ATAACTTACAGACAGGCTTCAAGTCTTCTGGAAGTCAAGGATTCAGAGTTCG 3290
3286 TCGGCTGTGCGACAGGAGTCTCTCCCTCCGTTCCGCGCGAGAGTGTTCATGATTCG 3345
3291 TCGGCTGTGCGACAGGAGTCTCTCCCTCCGTTCCGCGCGAGAGTGTTCATGATTCG 3350
3346 AATACGCTTACGAGCTCAACATGAGCAGCAAGCGGTGTGAGAGTTCATCTTTTACT 3405
3351 AGTATGATACCTACCTGAAACAAAGGAGTCAAGGAGTGTGAGAGTTCATCTTTTACT 3410

3406 GCGTGAATATTTCCCTTCCAGATGTGAGAACGGGCAACCACTTACCTTCAGCTACA 3465
3411 GCGTGAATATTTCCCTTCCAGATGTGAGAACGGGCAACCACTTACCTTCAGCTACA 3470
3466 CTTTGAAGAGTGTCTTCCAGAGCACTACGCGACAGCAGCAGCAGCAGCAGCAGCAGCAG 3525
3471 CTTTGAAGAGTGTCTTCCAGAGCACTACGCGACAGCAGCAGCAGCAGCAGCAGCAGCAG 3530
3526 TGAATCTCTCATTCGACCAATTACCTGTATTAAGTGAACAGAACTCAAAATCACTCCGGA 3585
3531 TGAATCTCTCATTCGACCAATTACCTGTATTAAGTGAACAGAACTCAAAATCACTCCGGA 3590
3586 GTGCGCAAAACAAGACTTCTGTAGCCGTGGTCTCCAGCTGACATCTGTTCAGC 3645
3591 CCACACAGCAGTCAAGCTTCAAGTCTTCTCAGCCGAGAGGAGTACATTCGGGACCACT 3650
3646 CCAAAAATGCGTCACTGAGCCCTGTATTCGAGACAGCGCGTCTTAAACAAACAG 3705
3651 CTAGAAATGCGTCTTCTGAGCCCTGTATTCGCGCAGCAGCAGAGTATCAAAAGCATTCGCG 3710
3706 ACAACAAACAAGCAATTTTACCTGAGCTGTGTCTTCAAAATTAACCTTCAATGGGCGTG 3765
3711 ATAAACAACAAGTGAATCTGTGAGCTGTGTCTTCAAAATTAACCTTCAATGGGCGTG 3770
3766 AATCATCATCAACCTGAGCACTGTCTTCAACCAAGAGCAGAGCAAGCAAGTCT 3825
3771 ACTCTGTGAGTCTTCCG 3830
3826 TTCCCATGAGCGGTGATATATTTTGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3885
3831 TTCCCATGAGCGGTGATATATTTTGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3890
3886 TGAACATGTCAATGATTAAG 3945
3891 TGAACATGTCAATGATTAAG 3950
3946 AAAGATTTGGGACCGGTGCGAGTCAATTTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAG 4005
3951 AGCAGTATGCTTCTGTATCTACCAACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4010
4006 ATGTGATGTATGAGAGCACTTACCTGAGAGTGTGGCAAGATGAGAGCTGTACTGC 4065
4011 ATGTGATGTATGAGAGCACTTACCTGAGAGTGTGGCAAGATGAGAGCTGTACTGC 4070
4066 AGGTCCTATTTGGGCAAAATTTCTCAACAGATGAGAGCACTTCCCGGCTCTCTTA 4125
4071 AGGTCCTATTTGGGCAAAATTTCTCAACAGATGAGAGCACTTCCCGGCTCTCTTA 4130
4126 TGGGCGCTTTGAGTCAAGAAACCGGCTCTCAGATCTCATCAAAACACGCGCTTTC 4185
4131 TGGGCGCTTTGAGTCAAGAAACCGGCTCTCAGATCTCATCAAAACACGCGCTTTC 4190
4186 CTGCAATCTCTCGCGGAGTCTTCAAGTCAAAAGTGTCTCATTCATCAACCAATACT 4245
4191 CTGCAATCTCTCGCGGAGTCTTCAAGTCAAAAGTGTGTCTCTCATTCATCAACCAATACT 4250
4246 CCACAGACAGAGTGTGAGAAATTTGAATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 4305
4251 CCACAGACAGAGTGTGAGAAATTTGAATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 4310
4306 GGAATCCGGAAGTCAAGTCAACATTCATCAAAATTTGCAACAGTGTATTTACTG 4365
4311 GGAATCCGGAATTCAGTCAACATTCATCAAAAGTGTATTTGAGACTTTACTG 4370
4366 TGAACAACAATGAGCTTATTAAGTGTGAGCTTCCAGCAGTGTACCTTAACTCCGTC 4425
4371 TGAACAACAATGAGCTTATTAAGTGTGAGCTTCCAGCAGTGTATTTGAGACTTTACTG 4430
4426 CCTGTATTAAGTGTATTAATCAATTAACCGGTGATTTGTTTCAAGTGAACCTTTGCTC 4485
4431 ATCTGTATTAAGTGTATTAATCAATTAACCGGTGATTTGTTTCAAGTGAACCTTTGCTC 4490

QY 4486 CTGCTCTTATCTATC 4504
DB 4491 TGGTATTCTTCTTATC 4509

RESULT 12
ID AAD00834 standard; DNA; 8178 BP.
XX AAD00834;
XX AAD00834;
XX 08-SEP-2000 (first entry)
XX
XX Chimeric adeno-associated virus 2/3 capsid encoding helper plasmid.
XX
XX Adeno-associated virus; AAV2; AAV3; chimeric; gene therapy; plasmid;
KW recombinant parvoviral vector; cellular tropism; rep gene; cap protein;
KM capsid; gene delivery; VP1; VP2; VP3; s8.
XX
XX Chimeric - Adeno associated virus serotype 2.
OS Chimeric - Adeno associated virus serotype 3.
XX
XX Key Location/Qualifiers
FT CDS 251..2180
FT /*tag= a
FT /product= "Rep 68"
FT /note= "The specification states that this region
FT encodes Rep 68 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 251..2114
FT /*tag= b
FT /product= "Rep 78"
FT /note= "The specification states that this region
FT encodes Rep 78 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 923..2114
FT /*tag= c
FT /product= "Rep 52"
FT /note= "The specification states that this region
FT encodes Rep 52 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 923..2180
FT /*tag= d
FT /product= "Rep 40"
FT /note= "The specification states that this region
FT encodes Rep 40 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 2133..4342
FT /*tag= e
FT /product= "VP1 cap protein"
FT /note= "The specification states that this region
FT encodes VP1 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 2544..4342
FT /*tag= f
FT /product= "VP2 cap protein"
FT /note= "The specification states that this region
FT encodes VP2 protein of AAV2. However, the positions
FT given in the specification for this CDS do not correspond
FT to start or stop codons"
FT 2739..4342
FT /*tag= g
FT /product= "VP3 cap protein"
FT /note= "The specification states that this region
FT encodes VP3 protein of AAV2 in which loops 2-4 are
FT replaced with the corresponding region from AAV3.
FT However, the stop position given in the specification
FT for this CDS does not correspond to stop codon"

FT misc_feature 3184..4092
FT /*tag= h
FT /note= "AAV3 sequence"
XX
XX WO200028004-A1.
XX
XX 18-MAY-2000.
XX
XX 10-NOV-1999; 99WO-US26505.
XX
XX 10-NOV-1998; 98US-0107840.
XX 10-MAR-1999; 99US-0123651.
XX
XX (UYN-) UNIV NORTH CAROLINA.
XX
XX Rabinowitz JE, Samulski RJ, Xiao W;
XX WPI; 2000-376523/32.
XX
XX Recombinant parvoviral vectors with altered packaging, tropisms and
PT immunogenic properties, useful in gene therapy protocols -
XX
XX Example 27; Page 143-147; 153pp; English.
XX
XX The patent discloses modified parvovirus vectors with advantageous
CC antigenic properties, packaging capabilities and cellular tropisms.
CC These vectors can be used in standard recombinant DNA protocols e.g. gene
CC therapy for delivering nucleic acids to cells. The present sequence is
CC a helper plasmid encoding a chimeric adeno-associated virus serotype 2/3
CC (AAV2/3) capsid. The sequence contains the AAV2 rep coding sequences,
CC most of the AAV2 capsid coding sequences with the exceptions that
CC loops 2-4 from the AAV2 VP3 subunit is replaced with the corresponding
CC region from AAV3, in a pluscript backbone. The rep
CC coding sequence encodes four proteins responsible for replication,
CC Rep 68, Rep 78, Rep 52 and Rep 40, and the capsid (cap) coding region
CC encodes three structural proteins VP1, VP2 and VP3.
CC Recombinant parvovirus comprising the chimeric capsid is useful
CC for gene delivery.
XX
XX Sequence 8178 BP; 2082 A; 2078 C; 2027 G; 1991 T; 0 other;
SQ

Query Match 62.4%; Score 2941.8; DB 21; Length 8178;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 3517; Conservative 0; Mismatches 792; Indels 20; Gaps 8;

QY 205 GTCTGTATTAGCTGTACGTGAGTCTTTTGGACATTTTGGACACCACTGSCCATT 264
DB 121 GTCTGTATTAGCTGTACGTGAGTCTTTTGGACATTTTGGACACCACTGTCACG 179
QY 265 TAGGATATATATGATCCGAGTGAAGGACGAGATCTTCATTTTG-ACCGGAATTGGAAC 323
DB 180 CTGGGTATTTTAAACCCCGAGTGAAGGACGAGGCTTCATTTTGAACGGGAGTTTGAAC 239
QY 324 GAGCAGACGACATGCGGAGCTTCTACGATCGTATCAAGTGGCGGAGCAGCTTGACG 383
DB 240 GCGCAGCCGCGCATGCGGGGTTTACGATTTGATTAAGTCCCGACGACTTGACG 299
QY 384 AGCAGCTGCGGGCATTTTCTGACTGTTTGTGAGCTGAGTGGCGGAGAAATGAGAGC 443
DB 300 GGCATCTGCCCGGCATTTCTGACAGCTTGTGAACTGGGTGGCGGAGAAATGAGAGT 359
QY 444 TGGCCCCGATTTGTGACATGATCTGAATCTGATTTAGACAGGACCCCTGACCTG 503
DB 360 TGGCGCAATTTGTGACATGATCTGAATCTGATTTAGACAGGACCCCTGACCTG 419
QY 504 AGAAGCTGACGCGCATCTCTGTGATGCGCGCGCTGAGTAAAGCCCGAGAGCC 563
DB 420 AGAAGCTGACGCGCATCTCTGTGATGCGCGCGCTGAGTAAAGCCCGAGAGCC 479
QY 564 TCTTCTTTGTTGATGAGAAAGGCGAGTCTTACTTCACTTCATTTCTGAGAGA 623
DB 480 TTTTCTTTGTTGATGAGAAAGGCGAGTCTTACTTCACTTCATTTCTGAGAGA 539

QY	624	CCACGGGGGTCAAAATCCATGTGTGCTGGGCGCTTCTGTGAAGTACATTTAAGGACAAAGCTGG	683
Db	540	CCACCGGGGGTGAATCCATGGTTTTTGGGACGTTTCTGTGAAGTACATTTGCGCAAAAACCTGA	599
QY	684	TGCAGACCATCTACCGCGGGGATTCGAGCGGACCCCTGSCCCCACTGGTTCGCGGTGACCAAGA	743
Db	600	TTTCAGAGATTTTACCGCGGGGATTCGAGCCGACTTTGCCAAATCTGGTTCGCGGTGACCAAGA	659
QY	744	CGCGTAATGCGCGCGGAGGGGGGAAACAAGTGTGTGACGAGTGTCTACATCCCAACTACC	803
Db	660	CCAGAAATGCGCGCGGAGGGGGGAAACAAGTGTGTGAGTGAAGTCTACATCCCAACTACTCT	719
QY	804	TCTGTCCCAAGACTTCAGGCCGAGCTGTGACGTGGGCGTGTGACTTCATGTAGAGAGATATATA	863
Db	720	TGCTCCCAAAACCCAGCGCTGTGAGCTCCAGTGGCGCTGTGACTTCATATGTAGAACAGTATTTAA	779
QY	864	GCGCGTGTTTGAACCTGGCGGAGCGCAACCGGCTGTGCGGAGACCTTGACCCACGCTCA	923
Db	780	GCGCTGTGTGATTTCAACGAGCGCTAAACGGTTGTGTGGTGGATTTGACGACGCTGT	839
QY	924	GCCAGACCCAGGAGCAGAACAAAGAGAAATCTGAACCCCAATTCGACCGCGCTGTCACTCC	983
Db	840	CGCAGACCCAGGAGCAGAACAAAGAGAAATCTGAATCCCAATTCGATGCGCGCGGTGATCA	899
QY	984	GGTCAAAAACCTCCGCGCGCTTCATGTAGAGCTGTGTGGTGGTGTGTGTGACCGCGGGATCA	1043
Db	900	GATCAAAAACCTTCACGCCAGAGTACATGAGAGCTGTGTGGTGGTGTGTGTGACCAAGGGATTA	959
QY	1044	CCTCCGAGAGCAGTGGATTCGAGGAGGACGAGGCGCTGTGACATCTCTTCAACGCGGCTT	1103
Db	960	CCTCGAGAGACAGTGGATTCGAGGAGGACGAGGCGCTTCATCATCTCTTCAATGCGGCTT	1019
QY	1104	CCAACTCGCGGTCCCAAGTCAAAGCGGCTGTGACAAATGCCGCAAGATCATGTGCGCTGA	1163
Db	1020	CCAACTCGCGGTCCCAAAATCAAAGCTGTGCTTGTGACAAATGCGGAAAGATTTAGAGCTGA	1079
QY	1164	CCAAATCCGCGCGGACTTACTGTGTAGGCGCCGCTCGCGCCGCGGACATTTAAACCAACC	1223
Db	1080	CTAAACCGCGCCCACTCACTGTGTGGGCGCAGACCGCGTGAAGACATTTCCAGCAATTC	1139
QY	1224	GCATTTACCGCATCTGTGAGGCTGAACCGGCTACGAACCTGCCATCGCGGCTCCGCTTTC	1283
Db	1140	GGATTTATTAATTTTGTGATCTAAACGGGTACGATCCCAATATCGGCTTCCGCTTTC	1199
QY	1284	TCGCGTGGGCCCCAGAAAAGTTTCGGGAAGCGCAACCATCTGTGCTGTTTGGCGCGGCA	1343
Db	1200	TGGGATGGGCGACGAAAAGTTTCGGGCAAGAGGAAACCATCTGTGCTGTTTGGGCGTGA	1259
QY	1344	CCACGGGCAAGACCAACATTCGCGGAAGCCATGCGCCACGCGCTGCTTCTTACCGGCTGG	1403
Db	1260	CTACCGGGAAGACCAACATTCGCGGAGGCGCATGCGCCACACTGTGCTTCTTACCGGCTGG	1319
QY	1404	TCAATCTGACCAATAGAACTTTCCTTCATAGATTGTGTGTGACAAAGATGTGATCTGGT	1463
Db	1320	TAAATCTGACCAATAGAACTTTCCTTCATAGACTGTGTGTGACAAAGATGTGATCTGGT	1379
QY	1464	GGAAGAGAGGCAAGATGACGCGCAAGTCTGTGATTCGCGCAAGGCGCATTTCTCGCGGCA	1523
Db	1380	GGAAGAGAGGGAAGATGACCGGCAAGTCTGTGATTCGCGCAAGGCGCATTTCTCGGAGGA	1439
QY	1524	GCAAGTGTGCGGTGACCAAAAATGTCAAGTGTCCGCGCCAGTGTGACCCACCCCGCTGA	1583
Db	1440	GCAAGTGTGCGGTGACCAAAAATGTCAAGTGTCCGCGCCAGTGTGACCCCGCTGTGA	1499
QY	1584	TGCTTACTCTCAACCAACATGTGTGCGCGTGTGATTTGAAGGGGAACGACCAACTTTCGAGC	1643
Db	1500	TGCTTACTCTCAACCAACATGTGTGCGCGGTGTGATTTGAAGGGGAACCTCAACGACTTTCGAC	1559
QY	1644	ACCAGCAGCGGTGTGAGAGCGGATGTTCAATTTGAACTCAACCGCGGTGTGAGCATG	1703
Db	1560	ACCAGCAGCGGTGTGAGAGCGGATGTTCAATTTGAACTCAACCGCGGTGTGATCATG	1619
QY	1704	ACTTTGGCAAGGTGACAAAGCAAGGCTCAAGAGTTCTTCCGTGGCGCAGAGTACG	1763

Db	1620	ACTTTGGGAAAGTACCAAGCAGGAAGTCAAAAGCTTTTCCGTTGGGCAAAAGATACG	1679
Qy	1764	TGACCGAGGTGGCGCATGAGTTCTACGTCAAAAAGGTGAGCCACAAAAGACCCGCC	1823
Db	1680	TGTTTGAAGTGGACATGAATTTCTACGTCAAAAAGGGTGGAGCCAAAGAAAAGCCGCC	1739
Qy	1824	CCGATGACGCGGATTAAGAAGGAGGCCCAAAGGGGCGTGGCCCTCAGTGCGGATTCATCGA	1883
Db	1740	CCAAGTCAGCAGATATTAAGTGAAGCCCAAAGGGGTGGCGAGTCACTTTGCCAGCATCGA	1799
Qy	1884	CGTCAGACGCGGAGAGAGCTCCGGTGAAGCTTTGGCCAGAGTACCAAAACAAATGTTCTC	1943
Db	1800	GCTCAGACGCGGGA---AGCTTCAATCACTTACGACGACAGTACCAAAACAAATGTTCTC	1856
Qy	1944	GTCACGCGGGCATGCTTCAAGTGTCTTTTCCCTGCAAGACATGCGAGAGATGAATCAGA	2003
Db	1857	GTCACGTTGGGATGAATCTGATGTCTTTCCCTGCAAGACATGCGAGAGATGAATCAGA	1916
Qy	2004	ATTTCAAATTTGCTTCAACGCAAGGACGAGAGACTTTCAAGAGTCTTCCCAGGCGTGT	2063
Db	1917	ATTCAAAATATCTGCTTCACTCAACGCAAGAAAGCTTTTAAGATGTCTTCC---CGTGT	1973
Qy	2064	CAGAAATCCAAACCGGT---GTCAGAAAGAGAGATATCGGAAATCTGTGCGCATTCATC	2120
Db	1974	CAGAAATCCAAACCGGTCTTCTGTCTCTCAAAAAGCGTATCAAGAACTGTCTACATTCATC	2033
Qy	2121	ATCTGCTGGGCGGGCTCCCGAGATTGCTCTGCGGCTGCGATCTGTCAACCTGGACC	2180
Db	2034	ATATCATGGGAAAGTGGCCAGA---CGCTTGCACTGCTGCGATCTGTCAATGTGATT	2090
Qy	2181	TGATGACTGTGTTCTTGAGCAATTAATGACTTAAACAGATATGCTGCCGATGTATT	2240
Db	2091	TGATGACTGTGATCTTTTGAACAAATTAATGATTTAAACAGTATAGGCTGCCAGTGTAT	2150
Qy	2241	CTTCAGATTGGCTCGAGGACCAACTCTCTGAGGGGCAATGCGGAGTGTGGGACTTGA	2300
Db	2151	CTTCAGATTGGCTCGAGGACCTCTCTTGAAAGAAATTAAGACAGTGTGGAGCTGAAA	2210
Qy	2301	CCTGAGGCCCCGAAAGCCCAAGCCCAACAGCAAGCAAGGAGCGAGCGGGTCTTGGTG	2360
Db	2211	CCTGGCCACCAACCAACAAAGCCCGCAGAGGGGCAATTAAGGACGACAGAGGGGCTTGTG	2270
Qy	2361	CTTCTCGGCTACAGATACCTCGGACCTCTTACCGAATCTCAAGAGGGGAGCCCTGTAAC	2420
Db	2271	CTTCTCGGCTACAGATACCTCGGACCTCTTACCGAATCTCAAGAGGGGAGCCCTGTAAC	2330
Qy	2421	GCGGCGGACCGAGGGGCGCTCGACACAGCAAGGCGCTACGACCGAGCTCAAGCGGGT	2480
Db	2331	GAGGCGGACCGGCGGCGCTCGACACAGCAAAACCTTACGACCGGAGCTGACAGCGGA	2380
Qy	2481	GACAAATCCGATCTCGCGGATTAACACAGCGGACCGCGAGTTTCAGAGCGCTCTGCAAGAA	2540
Db	2391	GACAAATCCGATCTCAAGTACAAACAGCGGACCGGAGTTTCAGAGCGCGCTTAAAGAA	2450
Qy	2541	GATACGTTTGGGGGCAACTCTCGGACGAGCACTTTCAGAGCGGAAAGAGGGTCTT	2600
Db	2451	GATACGTTTGGGGGCAACTCTCGGACGAGCACTTTCAGAGCGGAAAGAGGGTCTT	2510
Qy	2601	GAACCTCTCGGTCTGTGTTGAGGAAGGGCTTAAGCGGCTCTGGAAGAAACGTCGCGGTA	2660
Db	2511	GAACCTCTCGGCGCTGTGTTGAGGAACCTGTTAAGCGGCTCCGGGAAAAAGAGGCGGTA	2570
Qy	2661	GAGCAGTTCGACCAAGAGCGAGACTCTCTCTCGGACATCGGCAAGACAGGCGACAGCCC	2720
Db	2571	GAGCAGTTCCTGTGAGGACCAAGCTCTCTCTCGGGAACCGGAAGGGGCGACACAGCT	2630
Qy	2721	GCTAAAAAGAGACTCAATTTTGTGACATGCGACTCAAGTCAAGTCCCCGATCAACA	2780
Db	2631	GCAAGAAAAAGATTTGAATTTTGTGACATGCGAACCAAGACTCAATCACTGACCCCCAG	2690
Qy	2781	CCTCTCGGAGAACTCCAGCAACCCCGGCTGCTGTGGGACTTAATAAGCTTACAGGC	2840

Db 2691 CCTCTCGGAGAGCCAGCAGAGCCCTCTGCTGGGAACTAATACGATGCTACAGGC 2750
 QY 2841 GGTGGGCGACCAANTGGAGCAATAAGAGGGCGGACGAGGGGTAATGCTCTGGA 2900
 Db 2751 AGTGGCCGACCAATGGAGCAATAAGAGGGCGGACGAGGGGTAATGCTCTGGA 2810
 QY 2901 AATTGGCATTTGCAATTCAGATGCTGGGCGAGAGATCTATCAACACAGACCCGAC 2960
 Db 2811 AATTGGCATTTGCAATTCAGATGCTGGGCGAGAGATCTATCAACACAGACCCGAC 2870
 QY 2961 TGGGCTTTGGCCACTTCAATAACCACTCTCAAGAGAAATTCAGTCTTCAACGG 3020
 Db 2871 TGGGCTTTGGCCACTTCAATAACCACTCTCAAGAGAAATTCAGTCTTCAACGG 2927
 QY 3021 GCCAGCAAGCAACCACTCTCTGCTAGAGACCCCTGGGGGTAATTTGATTCAAC 3080
 Db 2928 GCTTCGAGCAATCACTCTTGGCTAGAGACCCCTGGGGGTAATTTGATTCAAC 2987
 QY 3081 AGATTCACCTGCACTTTTCAACAGCTGAGCGAGCACTCATCAACAATTTGGGGA 3140
 Db 2988 AGATTCACCTGCACTTTTCAACAGCTGAGCGAGCACTCATCAACAATTTGGGGA 3047
 QY 3141 TTCGGGCTCAAGAGACTCAACTTCAAACTTTCAACATTCAGTCAAGAGAGTCAAG 3200
 Db 3048 TTCGAGCCCAAGAGACTCAACTTCAAACTTTCAACATTCAGTCAAGAGAGTCAAG 3107
 QY 3201 AATGATGGCGTCAACCACTCTGCTAATACCTTACAGAGCGGTCAAGTCTTCTGGAC 3260
 Db 3108 AATGATGGCGTCAACCACTCTGCTAATACCTTACAGAGCGGTCAAGTCTTCTGGAC 3167
 QY 3261 TCGGAGTACAGCTTCCGTAAGTCTCTGGCTGCGCAGCAGGAGGCTCTCCCTCCGTT 3320
 Db 3168 TCGGAGTACAGCTTCCGTAAGTCTCTGGCTGCGCAGCAGGAGGCTCTCCCTCCGTT 3227
 QY 3321 CCGGCGGAGCTGTTCAATGATTCGCAATACGCTGCTGACGCTCAACAAATGGAGCAA 3380
 Db 3228 CCGGCGGAGCTGTTCAATGATTCGCTCAATGATGATCACTCAACAAATGGAGCAA 3287
 QY 3381 GCGGTGGAGAGTTCATCTTTTACTGCTGCTGAAATTTCCCTCTGAGATGCTGAGAAC 3440
 Db 3288 GCGGTGGAGAGTTCATCTTTTACTGCTGCTGAAATTTCCCTCTGAGATGCTGAGAAC 3347
 QY 3441 GCGCAACAACTTTTACTGCTGCTGAGATGCTGCTTTCACAGAGCTAGCG 3500
 Db 3348 GCGCAACAACTTTTACTGCTGCTGAGATGCTGCTTTCACAGAGCTAGCG 3407
 QY 3501 CACAGCCAGAGCTGAGCGGCTGATGATCTCTCATGCAACCAATCACTGTATTAC 3560
 Db 3408 CACAGCCAGAGCTGATGCTGCTGATGATCTCTCATGATGATGATGATGATGATG 3467
 QY 3561 AAGAGAACTCA---AATCACTGCGGAGAGTCCCAAAACAAGAGACTTGTCTTAC 3617
 Db 3468 AAGAGAACTCA---AATCACTGCGGAGAGTCCCAAAACAAGAGACTTGTCTTAC 3527
 QY 3618 GGGTCTTCAGACTGAGTGTCTGTTCAAGCCCAAAAATGCTGATCCCTGTATTCCG 3677
 Db 3528 GGGTCTTCAGACTGAGTGTCTGTTCAAGCCCAAAAATGCTGATCCCTGTATTCCG 3587
 QY 3678 CAGCAGCGCTTTCTTAAACAAAACAGACCAACCAACGCAATTTTACCTGAGCTGT 3737
 Db 3588 CAGCAGCGCTTTCTTAAACAAAACAGACCAACCAACGCAATTTTACCTGAGCTGT 3647
 QY 3738 GCTTCAAAATATATACCTCATATGGCGGTGATTCATCATCTGCACTGTGATAGCC 3797
 Db 3648 GCTTCAAAATATATACCTCATATGGCGGTGATTCATCATCTGCACTGTGATAGCC 3707
 QY 3798 TCACACAAAGACAGAGCAAGATTTCTTCCCATGAGCGGTGATTTTGGGAAA 3857
 Db 3708 AGTCACAAAGACAGAGCAAGATTTTCTTCCCATGAGCGCAATCTAATTTTGGGAAA 3767
 QY 3858 GAGAGCGCCGAGAGTTCAAACACTGATGAGCAATGCTATGATTTACAGAGAGGAA 3917
 Db 3768 GAGAGCGCCGAGAGTTCAAACACTGATGAGCAATGATGATTTACAGAGAGGAG 3827

QY 3918 ATTAAAGCACTTAACCTGTGGCCACCGAAAGATTGGACCGTGGCAGTCAATTTCCAG 3977
 Db 3828 ATTGTAACCACTAATCTGTGGCCACCGAAAGATTGGACCGTGGCAGTCAATTTCCAG 3887
 QY 3978 AGCAGAGCAACAGACCTGTGGCCACCGAAAGATTGGACCGTGGCAGTCAATTTCCAG 4037
 Db 3888 AGTCAAAATACAGCTCCAGACTGAGTCAATCAATCAAGAGGAGCTTACCTGGCATG 3947
 QY 4038 GTGTGCAAGATGAGAGCTGTACCTGCAAGGCTCCATTTGGGCAAAATTTCTCAACA 4097
 Db 3948 GTGTGCAAGATGAGAGCTGTACCTGCAAGGCTCCATTTGGGCAAAATTTCTCAACA 4007
 QY 4098 GATGACACTTTGACCGGTCTCTTATGGGGGCTTTGGAGTCAAGAACCCGCTCTCT 4157
 Db 4008 GATGACACTTTGACCGGTCTCTTATGGGGGCTTTGGAGTCAAGAACCCGCTCTCT 4067
 QY 4158 CAGATCTCATCAAAACAGCGCTGTCTGCGAATCTCCGCGAGATTTTCAGCTACA 4217
 Db 4068 CAGATCTCATCAAAACAGCGCTGTCTGCGAATCTCCGCGAGATTTTCAGCTACA 4127
 QY 4218 AAGTTGCTTCAATTCATCAACCAATCTCCAGAGCAAGATGATGGAATGGAATG 4277
 Db 4128 AAGTTGCTTCAATTCATCAACCAATCTCCAGAGCAAGATGATGGAATGGAATG 4187
 QY 4278 GAGCTGCAAGAAAGAAACAGAGCGCTGGAATCCCGAAGTGCATGATCAATCAAT 4337
 Db 4188 GAGCTGCAAGAAAGAAACAGAGCGCTGGAATCCCGAAGTGCATGATCAATCAAT 4247
 QY 4338 GCAAAATCTGCCAACCTGATTTTACTGTGAGCAACAAATGAGACTTATCTAGAGCTCC 4397
 Db 4248 AACAATCTGTTAATGATGAGACTTACCGTGAATCAATGAGCTGATTCAGAGCTCC 4307
 QY 4398 CCCATTGGACCCGTTAATCTTACCGCTCCCTGTAAATGATGATGATGATGATGATG 4457
 Db 4308 CCCATTGGACCCGTTAATCTTACCGCTCCCTGTAAATGATGATGATGATGATGATG 4367
 QY 4458 TGAATGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4517
 Db 4368 TGAATGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4427
 QY 4518 ATAGCTTAC 4526
 Db 4428 CTGAGCTAC 4436

RESULT 13
 AAD00832
 ID AAD00832 standard; DNA; 8151 BP.
 AC AAD00832;
 XX
 DT 08-SEP-2000 (first entry)
 XX
 DE Chimeric adeno-associated virus 2/4 capsid encoding helper plasmid.
 XX
 KW Adeno-associated virus; AAV2; AAV4; chimeric; gene therapy; plasmid;
 KW recombinant parvoviral vector; cellular tropism; rep gene; cap protein;
 KW capsid; gene delivery; VP1; VP2; VP3; ss.
 OS Chimeric - Adeno associated virus serotype 2.
 OS Chimeric - Adeno associated virus serotype 4.
 FH Location/Qualifiers
 FT CDS
 FT 251..2180
 FT /tag= a
 FT /product= "Rep 68"
 FT /note= "The specification states that this region
 FT encodes Rep 68 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 251..2114
 FT CDS
 FT /tag= b

QY	1164	CCAAATCCGCGCCGACATACCTGTAAGGACCCCGCTCCGCGCGGACATTTAAACCAAC	1223
Db	1080	CTAAACCCGCCCTCCATCTTGGTGGCCAGCACCCCTGTGAGAGACATTTCCAGCAATC	1139
.QY	1224	GCATCTACCGGATCCTGGAGCTGAACGGCTACGAACCTGACCTAGCCGGCTCCGTCTTTC	1283
Db	1140	GGATTTATTAATAATTTTGGACCTAAACGGGTAGCATCCCAATATAGCGGCTTC	1139
QY	1284	TCGGCTGGGCCGAGAAAAGTTTCGGGAGCCGACACATCTTGGCTGTTTGGGCCGACCA	1343
Db	1200	TGGGATGGGCCACGAAAAAGTTTCGGCAAGAGAACCATCTGGCTGTTTGGGCTGCAG	1259
QY	1344	CCAGGGGCAAGACCAACATCCGCGAAGCAATCCGCCACGCCGTCCTTACACGGTTCG	1403
Db	1260	CTACCGGGAGAACCAACATCCGCGAGGCGCATAGCCACATGTGCTTCTACGAGTTCG	1319
QY	1404	TCAACTGACCAATAGAGATCTTCCCTTCAATGATGTGCTGACAGACAGATGTGATCTGGT	1463
Db	1320	TAAACTGGAACCAATGAGAACCTTCCCTTCAAGACTGTGTGACAGATGTGATCTGGT	1379
QY	1464	GGAGAGAGGGCAAGATGACGGCCAAAGTCGTGAGTCCGCCAAGGCCATTTCTCGCGGCA	1523
Db	1380	GGGAGAGAGGAAAGATGACCCGCAAGTCGTGAGTCCGCCAAGGCCATTTCTCGAGGAA	1439
QY	1524	GCAAGGTCCGGGTGGACCAAAAGTGCATGTGTCGCCCAATGACCCCCACCCCCGTGA	1583
Db	1440	GCAAGGTCCGGGTGGACCAAAAGTGCATGTGTCGCCCAATGACCCCCACCCCCGTGA	1499
QY	1584	TGCTCACCCTCCAAACCAACATGTGCGCGCGTATGACGGGAAACGACACCACTTTCGAGC	1643
Db	1500	TCGTCACTCCAAACCAACATGTGCGCGCGTATGACGGGAAACGACACCACTTTCGAGC	1559
QY	1644	ACAGCAGCCGTTGCAGGACCCGATGTTCAATTTGAACTGACCCGCCGTCTGGAGCATG	1703
Db	1560	ACAGCAGCCGTTGCAGGACCCGATGTTCAATTTGAACTGACCCGCCGTCTGGAGCATG	1619
QY	1704	ACTTTGGCAAGTGCACAAAGCAGAGATCTTTCGCTGGCGCAGAGATCAAG	1763
Db	1620	ACTTTGGCAAGTGCACAAAGCAGAGATCTTTCGCTGGCGCAGAGATCAAG	1679
QY	1764	TGACCGAGGTGAGCGCATGATGTTCTACGTACAGAAAGGTTGAGACCAACAAAGACCCGCC	1823
Db	1680	TGTTTGGAGTGGAGATTAATTTCTACGTACAAAGAGTGGAGCCAGAAAGACCCGCC	1739
QY	1824	CCGATGACGCGGATTAAGACGAGCCCAAGCGGCTGCCTCACTAGTCGCGATCCATCGA	1883
Db	1740	CCAGTGAAGCAGATTAATTAAGTGAGCCCAAGCGGTGCGGATCGATGTCGAGCATCGA	1799
QY	1884	CGTCAGACGCGGAGAGAGCTCCGGTGAACCTTCCGACAGGTATCCAAACAAATGTTCTC	1943
Db	1800	CGTCAGACGCGGAGAGAGCTCCGGTGAACCTTCCGACAGGTATCCAAACAAATGTTCTC	1856
QY	1944	GTCACGCGGGGATGCTCAGATGCTGTTCCCTGCAACACATGTGAGAGATGAATCAGA	2003
Db	1857	GTCACGTTGGGCAATGATTAATGCTGTTCCCTGCAACATGTGAGAGATGAATCAGA	1916
QY	2004	ATTTCAACATTTGCTTCAACGACGAGACCTGTTCAAGTGTCTCCCGCGCGTGT	2063
Db	1917	ATTTCAACATTTGCTTCAACGACGAGACCTGTTCAAGTGTCTCCCGCGCGTGT	1973
QY	2064	CAGAACTCAACCGGT---CGTCAGAAAGAGAGCTATCGGAAACTGTGCTCATCTAC	2120
Db	1974	CAGAACTCAACCGGT---CGTCAGAAAGAGAGCTATCGGAAACTGTGCTCATCTAC	2033
QY	2121	ATTCGCTGGGGGCGGCTCCGAGATGTGTCGTGGGCTGGAGATCTGTGCAACGTGACC	2180
Db	2034	ATTCATGTGGAAAGGTGCGAG---CGCTTGACATGCTCTGAGATCTGTGCAATGTGAAT	2090
QY	2181	TGAGATGATCTGTTTCTGAGCAATTAAGACTTAACAGGATATGCTGCGAGATGTAAT	2240
Db	2091	TGAGATGATCTGATCTTTTGAACATTAAGATTTTAATCAGGATATGCTGCGAGATGTAAT	2150
QY	2241	CTTCAGATTTGGCTGAGAGCACTCTCTGAGGCAATTCGCGAGTGTGGACTTGAA	2300

[illegible]

```

Db 3204 CCAGCAGACGCTTCATGATGGCCACAGATGATACCTCACCTTGAAACAAGGAGTCAG 3263
Qy 3381 GCGGTGGAGGTTGATCCTTTTATCTGCTGGAAATATTTCCCTTTGAGATGCTGAAGAC 3440
Db 3264 GCAGTAGAGCGCTCTTCATTTTACCTGCTGGAGTACTTTCCTTCTCAGATGCTGCGTACC 3323
Qy 3441 GGCAACAACCTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTCCACAGAGCTACGCG 3500
Db 3324 GGAAACAACCTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTCCACAGAGCTACGCT 3383
Qy 3501 CACAGCAGAGCTGGAACCGGCTGATGATCTCTCAATCGAACCAATACCTGTATTACCTG 3560
Db 3384 CACAGCAGAGCTGGAACCGGCTCATGATCTCTCAATCGAACCAATACCTGTATTACCTG 3443
Qy 3561 AACGAACTCAAATTCAGTCCGGAAGTCCCAAAACAAGACTTGTCTTTAGCCGTGGG 3620
Db 3444 AGCAGAAACAACCTCCAGTGAACCAACGAGTCAAGCTTCAGTTTCTCAGGCC 3503
Qy 3621 TCTCAGCTGGCATGTCTGTTCAAGCCCAAAAATGAGTACCTTGACCTGTTATCGCAG 3680
Db 3504 GAGCGAGTGCATTTCCGGACCACTGAGAACTGCTCTCTGACCTCTGTTACCGCCAG 3563
Qy 3681 CAGCGCTTTCTAAAACAAAACAACAACAACAAGCAATTTTACTGCTGACTGTGCT 3740
Db 3564 CAGCGAGTATCAAAAGACATCTGCGGATTAACAACAAGTGAATCTGCTGAGTGAAGCT 3623
Qy 3741 TCMAAATTTAATCTCAATGGGCGTGAATCCATCAATCAACCTTGACCTGCTATGGCCCA 3800
Db 3624 ACCAAGTACCACTCAATGAGCAAGACTCTGTGGAATCCGAGCCGCGCATGCAAGC 3683
Qy 3801 CACAAGAGCGAGAAACAAGTCTTTCCATGAGCGGTGTCATGATTTTGGAAAAG 3860
Db 3684 CACAAGAGCGATGAAGAAATTTTCTCTGAGCGGGTTCATCTTTGGGAAGAA 3743
Qy 3861 AGCGCCGAGCTTCAACAACCTGCAATGCAATGTCAATTAACAAGAAAGAAAT 3920
Db 3744 GGCTCAGAGAAAACAATGTGAACATTGAAAAGTCAATGATTAACAAGAAAGAAATC 3803
Qy 3921 AAAGCCACTAACCTGTGGCCACCGAAGATTTGGGACCGTGGAGTCAATTTCCAGAGC 3980
Db 3804 GGAAACAACCAATCCCGTGGCTACGAGAGTATGTTGTATCTACCAACCTCCAGAGA 3863
Qy 3981 AGCAGCAGAGCCCTGACACCGAGATGTCATGCTATGAGAGATTACTGGAGATGTTG 4040
Db 3864 GGCACAAGACAGAGACTACCGAGATGTCAACACAGAGCGTTCTCCAGAGATGTTG 3923
Qy 4041 TGGCAAGATTAAGAGATGTCATGCAAGGCTCCATTTGGGCGAAATTTCTCAACAGAT 4100
Db 3924 TGGCAGGACAGAGATGTATCTTCAAGGCGCCCATCTGGGCAAGATTTCCACACGAGC 3983
Qy 4101 GGACACTTTCACCGCTCTCTTATATGGCGGCTTTGAGACTCAAGAACCGGCTCTCAG 4160
Db 3984 GGACATTTTCACCGCTCTCTCCTCATGGGTGATTCGACTTAAACACCTCTCTCCACAG 4043
Qy 4161 ATCCCTCATCAAAAACACGCTGTCTCTGCGAATCTCTCCGCGAGTTTTCAGCTACAAG 4220
Db 4044 ATTCTCATCAAGAAACACCGCTTACCTGGAATCTTTCGACCACTTCAGTGGCGCAAG 4103
Qy 4221 TTTGCTTATTCATCATCCCAATATCTCCACAGAGCAAGTGAATGGAATGGAATGGAG 4280
Db 4104 TTTGCTTCTTCATCATCAAGTACTCCAGGAGCAGGTCAAGTGAATGAGTGGAG 4163
Qy 4281 CTGCAAGAAAGAAACAGAGAGCGTGAATCCCGAATTCATCAATCAATATATCA 4340
Db 4164 CTGCAAGAAAGAAACAGAGAGCGTGAATCCCGAATTCATCAATCAATATATCA 4223
Qy 4341 AAATTCGCAACGTTGATTTTATCTGGAACAACAATGACTTTATATCTAGAGCTGCC 4400
Db 4224 AAGTGTGTAATCGTGAACCTTACCGTGAATCTAATGCGTGTATTCAGAGCTGCC 4283
Qy 4401 ATTGGACCCGTTACCTTACCCGCTGTAATTAAGTGTATCAATCAATCAACGGTTGA 4460
Db 4284 ATTGGACCAAGATCCTGATCTGTAATCTGTATGCTGTATTAATCAATCAACCGTTTA 4343

```

```

Qy 4461 TTGTTTCAAGTTGAACCTTGTGTCCTGCTCTTATCTTAACGTTACCAATGTTATA 4520
Db 4344 TTGTTTCAAGTTGAACCTTGTGTCCTGCTCTTATCTTAACGTTACCAATGTTATA 4403
Qy 4521 GCTTAC 4526
Db 4404 GACTAC 4409

RESULT 14
AAV21648
AAV21648 standard; DNA; 4767 BP.
AAV21648;
28-AUG-1998 (first entry)
AAV4 genome.
AAV: AAV-4; Rep gene; VP1 gene; capsid protein; vector;
gene transfer; gene delivery; cancer; gene therapy; ds.
Adeno associated virus 4.
Key Location/Qualifiers
promoter 130..291
/*tag= a
/*note= "p5"
372..2243
/*tag= b
/*product= Rep protein
2260..4467
/*tag= c
/*product= capsid protein
MO9811244-A2.
19-MAR-1998.
11-SEP-1997; 97NO-US16266.
11-SEP-1996; 96US-0025934.
(USSH ) US DEPT HEALTH & HUMAN SERVICES.
Chiorini JA, Kotin RM, Safer B;
WPI; 1998-207403/18.
P-PSDB; AAM46307-08.
Adeno-associated virus 4 based vectors - used to transduce erythroid
progenitor cells for treatment of e.g. cancer and other genetic
diseases
Claim 12; Page 46-49; 80pp; English.
This nucleotide sequence comprises the adeno associated virus 4
(AAV4) genome that includes open reading frames for the Rep protein
(see AAM46307) and the VP1 capsid protein (see AAM46308). The genome
was cloned from viral lysates amplified from COS and HeLa cells.
The invention provides recombinant vectors and viral particles based
on AAV4 that may be useful for transducing erythroid progenitor cells
e.g. for the treatment of cancer and genetic diseases which can be
corrected by bone marrow transplants using matched donors. A
claimed method of delivering a nucleic acid to a subject (including
a subject with antibodies to AAV2) comprises administering to a
cell from the subject an AAV4 particle comprising the nucleic acid
inserted between a pair of AAV inverted terminal repeats (see
AAV21651 and AAV21659), and returning the cell to the subject. Also
claimed is a method of screening for infectivity by AAV4 comprising
contacting the cell with AAV4 and detecting the presence of AAV4 in
the cells. Cells containing the AAV4 genome, ORI, OR4 or

```

CC Individual Rep protein and capsid proteins genes are provided.
 XX Sequence 4767 BP; 1137 A; 1404 C; 1333 G; 891 T; 2 other;

Query Match 55.0%; Score 2596; DB 19; Length 4767;
 Best Local Similarity 74.5%; Pred. No. 0;
 Matches 3582; Conservative 0; Mismatches 1096; Indels 129; Gaps 20;

QY 1 TTGCCCCCTCCCTCTGCGCGCTGCGCTGCGTGGGGGCGCTCCGAGCAAGATGTCGC 60
 DB 1 TTGGCCACTCCCTCTATGCGCGCTGCGCTCACTGCGCTGGAACCAAGGTCCTCC 60
 QY 61 AGACGAGAGAGCTGCTGCTGCGCGCCCAACGAGCAGCGCGCGCAGAGAGAGT 120
 DB 61 AGACTGCGCGCTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGT 120
 QY 121 GCGAATCTCACTAGAGG-----GTAAATCGGAAGCGCTCCAC 161
 DB 121 GCGAATCTCACTAGAGGTTTGCCCACTGAGTCAATGAGAGTCTAGGGTTAGGAG 180
 QY 162 GCTGCGCGCTGAGCGTGAATTAATGATAGAGAGAGTGTCTCTA-----T 213
 DB 181 GTCCCTATTAAGAGTACGTAGTGTCTATTTGCGAGCGTAGCGGAGCGATACC 240
 QY 214 TAGCTGACAGTGA-----GTGCTTTTGAGATTTTGGACACCACTGCGCAT 263
 DB 241 AAGCTGCAAGTCAACGACAGAGTGTGCTTTGGAGAGTTTGGACACCACTGCGCAT 300
 QY 264 TTAAGGTATATATGCGCGAGTGAAGC-AGATCTCAATTTTGAACCGGAAATTTGAA 322
 DB 301 G-AGGGTATATACCGCGAGTGAAGCAGCGAGAGTCAATTTTGGCGGAAATTTGAA 359
 QY 323 CGAGCAGAGCAGTCCGCGCTTCTAGAGATGCTGATCAAGGAGCGGAGCGAGCTGAG 382
 DB 360 CGAGCAGAGCAGTCCGCGGTTCTAGAGATGCTGATCAAGGAGCGGAGCGAGCTGAG 419
 QY 383 GAGCAGCTGCGCGAGTCTGACTGCTTTGAGAGTGGTGGCGGAGAGAGATGGAG 442
 DB 420 GAGCAGCTGCGCGAGTCTGACTGCTTTGAGAGTGGTGGCGGAGAGAGATGGAG 479
 QY 443 CTGCCCCCGAGTTTGAATGATGATCTGAATCTGAATTGAGAGAGCACTTGAAGTGG 502
 DB 480 CTGCCCCCGAGTTTGAATGATGATCTGAATCTGAATTGAGAGAGCACTTGAAGTGG 539
 QY 503 GAGAGCTGAGAGCGGCTTCTGAGTCCAAATGAGCGCGGAGTGAAGAGCGCGAGAGCC 562
 DB 540 GAGAGCTGAGAGCGGCTTCTGAGTCCAAATGAGCGCGGAGTGAAGAGCGCGAGAGCC 599
 QY 563 CTCTTCTTTGTCAGTTCGAGAGAGGCGAGTCTTCACTTCCACTCATATTTCTGTGAG 622
 DB 600 CTCTTCTTTGTCAGTTCGAGAGAGGCGAGTCTTCACTTCCACTCATATTTCTGTGAG 659
 QY 623 ACCAGCGGAGTCAATTCATGATGCTGAGCGCTTCTGAGTCAATGAGTGAAGAGAGCTG 682
 DB 660 ACCAGCGGAGTCAATTCATGATGCTGAGCGCTTCTGAGTCAATGAGTGAAGAGAGCTG 719
 QY 683 GTGAGAGCAATTCAGCGCGGAGTGAAGCGAGCTTCTGAGTCAATGAGTGAAGAGAGCTG 742
 DB 720 GTGAGAGCAATTCAGCGCGGAGTGAAGCGAGCTTCTGAGTCAATGAGTGAAGAGAGCTG 779
 QY 743 ACCGATATGAGCGCGAGAGGAGGAGCAAGTGTGAGAGAGTGTCAATGAGTGAAGAGAGCTG 802
 DB 780 ACCGATATGAGCGCGAGAGGAGGAGCAAGTGTGAGAGAGTGTCAATGAGTGAAGAGAGCTG 839
 QY 803 CTCTGCGCAAGTCAAGCGCGAGCTGAGTGGCGTGAATCAATGAGAGAGATATA 862
 DB 840 CTCTGCGCAAGTCAAGCGCGAGCTGAGTGGCGTGAATCAATGAGAGAGATATA 899
 QY 863 AGGCGCTGTTGAACCTGCGCGAGCGCAACGCTGCTGCGAGAGAGCTGAGAGAGCTG 922
 DB 900 AGGCGCTGTTGAACCTGCGCGAGCGCAACGCTGCTGCGAGAGAGCTGAGAGAGCTG 959
 QY 923 AGCCAGACCAAGAGAGAGCAAGAGATCTGAATCCCAATTTCTGAGCGCGCTGATC 982

DB 960 TCGCAGCGCAG 1019
 QY 983 CGGTCAAAACCTCCGCGGCTATAGAGAGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1042
 DB 1020 AGGTCAAAACCTCCGCGGCTATAGAGAGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGT 1079
 QY 1043 ACCTCGAG 1102
 DB 1080 ACCTCGAG 1139
 QY 1103 TCCAGTCTGCGTCTCCAGATCAAGAGCGCTTGTGAATATGCGCGCAAGATCATGAGCTG 1162
 DB 1140 TCCAGTCTGCGTCTCCAGATCAAGAGCGCGCTTGTGAATATGCGCGCAAGATCATGAGCTG 1199
 QY 1163 ACCAAATCGCGCGGCTATAG 1222
 DB 1200 ACCAAATCGCGCGGCTATAG 1259
 QY 1223 CGCATCAAG 1282
 DB 1260 CGCATCAAG 1319
 QY 1283 CTGCGTGGGCGCAG 1342
 DB 1320 CTGCGTGGGCGCAG 1379
 QY 1343 ACCAGCGGAG 1402
 DB 1380 ACCAGCGGAG 1439
 QY 1403 GTCACTGAG 1462
 DB 1440 GTCACTGAG 1499
 QY 1463 TGGAG 1522
 DB 1500 TGGAG 1559
 QY 1523 AGCAAGTGGCGGTGAG 1582
 DB 1560 AGCAAGTGGCGGTGAG 1619
 QY 1583 ATGTGACCTTCAAG 1642
 DB 1620 ATGTGACCTTCAAG 1679
 QY 1643 CACGAG 1702
 DB 1680 CACGAG 1739
 QY 1703 GACTTTGAG 1762
 DB 1740 GACTTTGAG 1799
 QY 1763 GTGAG 1822
 DB 1800 GTGAG 1859
 QY 1823 CCCGATGAG 1882
 DB 1860 CCCGATGAG 1919
 QY 1883 AGCTGAG 1942
 DB 1920 AGCTGAG 1976
 QY 1943 CGTCAAG 2002
 DB 1977 CGTCAAG 2036
 QY 2003 AATTTCATTTGCTTCAAG 2062

2037 AATGAGACATTTGCTTCACGACGAGGCTCATGACCTGCGGAGTCTCC---CGTG 2093
2063 TCAGAAATCTCAAC-----CGGTGCTCAAGAAAGAGACATTCGAGAACTCTGCGCAT 2116
2094 TCAGAAATCTCAACCGGTCTGTGTCTGAGAAAGGAGAGATCAAGAACTGTGTCGAT 2153
2117 CATCATCTGAGGAGCGGCTCCGAGATTCCTGCTGCGCTGCGATCTGTCAACG 2176
2154 CATCATCATGAGGAGGCGCGAGGTGCTGCTGCGCTGCGATCTGCGCAATG 2213
2177 GACCTGATGATGTGTTCTGAGCAATTAATGACTTAACCAAGATGCTGCCATG 2236
2214 GACTTGATGATGATGATGAAATTAATGATCAACCAAGATGATCT---GACGG 2270
2237 TTATCTTCAGATTTGCTGAGGACCACTCTGAGGAGCATGCGGATGCTGAGGACT 2296
2271 TTACCTTCAGATTTGCTGAGGACCACTCTGAGGAGCATGCGGATGCTGAGGACT 2330
2297 GAAACCTGAGACCCCGAAGCCCAAGCAACGAGCAAGGAGGAGGAGGAGGAGGAG 2356
2331 GAAACCTGAGACCCCGAAGCCCAAGCAACGAGCAAGGAGGAGGAGGAGGAGGAG 2390
2357 GGTGCTTCCTGCTCAAGATTAATCTGAGACCTTCAAGGATCTCAAGGAGGAGGAG 2416
2391 TGTCTTCCTGCTCAAGATTAATCTGAGACCTTCAAGGATCTCAAGGAGGAGGAG 2450
2417 CAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2476
2451 CAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2510
2477 GGTGAGCAATTCCTGAGCTGCGGATTAATCAACGAGGAGGAGGAGGAGGAGGAG 2536
2511 CGGTGAGCAATTCCTGAGCTGCGGATTAATCAACGAGGAGGAGGAGGAGGAGGAG 2570
2537 AGAAGATTCCTGCTTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2596
2571 GGGGAGCAACCTGCTTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2630
2597 TCTGAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2656
2631 TCTGAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2690
2657 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2716
2691 GTTATGTAATTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2750
2717 GCGGCTTAAGAGAGACTCAATTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2776
2751 GCGGCTTAAGAGAGACTCAATTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2801
2777 ACAACCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2836
2802 ACCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2855
2837 AGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2896
2856 AGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2915
2897 AGGAAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 2956
2916 GGGTGAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 2975
2957 CACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3016
2976 AACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3024
3017 GGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3076
3025 -GAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3083
3077 CAACGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3136
3084 CAACGCTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3143

3137 GGGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3196
3144 GGGATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3203
3197 GAGCAATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3256
3204 GAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3263
3257 GAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3316
3264 GAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3323
3317 GTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3367
3324 TTTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3383
3368 CAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3427
3384 CACTTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3443
3428 GATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3487
3444 GATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3503
3488 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3547
3504 CTGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3563
3548 CTTGATTAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3604
3564 CTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3623
3605 GCTGTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3664
3624 CAATTAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3683
3665 ACCCTGTTATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3724
3684 GCTTCAATTAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3743
3725 TACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3769
3744 CACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3803
3770 CATCATTAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3829
3804 TGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3860
3830 CATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3889
3861 CAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3920
3890 CAATGCTATTAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3949
3921 GACTGCTATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 3980
3950 ATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4009
3981 GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4040
4010 GATGCTATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4069
4041 GACAGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4100
4070 TCCCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4129
4101 TCCCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4160
4130 CGGCTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4189
4161 TGCGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4220

QY	4190	GAATTCCTCCGGCGAGTTTTCAGCTACAAAGTTTCTTCATTCATCAACCAATCTCCAC	424.9
Db	4221	GAATTCCTCAACGACCTTCAGCTCTACTCCGGTAACCTCTTCATTCATCTCAGTACAGAC	428.8
QY	4250	AGGACAACTGAGTGTGAAATTGAATGGAGCTGCAGAAAGAAAACAGCAACGCTGGAA	430.9
Db	4281	TGGCAGAGTGTCCGGTGAGATTGAATGGAGATTCAGAAAGAGCGGTCCAAAACGCTGGAA	434.0
QY	4310	TCCGGAATGCAGTACACATCCAAATTATGCAAAATCTGCCAAAGTTTATTTTACTGTGGA	436.5
Db	4341	CCCGGAGGTCCAGTTTACTCTCAACTACGACAGCAAAACCTCTGTTGTGGGCTCCCGA	440.0
QY	4370	CAACAATGACCTTTATTACTGAGCCCTGCCCAATTGGCAACCGTTAACCTTACCCGTCCT	442.9
Db	4401	TCCGCTGGGAAATACACTGAGCCTAGGGCTATTCGGTACCCGCTACACCCACCACTT	446.0
QY	4430	GTAATTACGTGTATCAATCAATAAACCGGTGTAATTCCTTCAGTTGAACTTTGGTCTCC-TG	448.8
Db	4461	GTAATTACCTTTATCAATCAATAAACCGGTTTATTCGTTTCAGTTGAACCTTTGGTCTCCGTG	452.0
QY	4489	TCCTTCTTATCTTATC-GGTTACCAATGTTATAGCTTAACAATTAAC TG-----CTTGG	454.1
Db	4521	TCCTTCTTATCTTATCTGTTTCCATGAGCTACTGCGTACATTAAGCAGCGGCTCGCGCGC	458.0
QY	4542	TTTGGGCTTCGGATTAAGACTTCG-----TCATGGGGTTACCCCTAGTATGG	459.1
Db	4581	TTTGGGCTTCGGGTTTACACTGCGCGGTTTATCACTAACTTCTGGCAAAACAGATGAATGG	464.0
QY	4592	AGTTGCCACATCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGGACCAAAAGTCC	465.1
Db	4641	AGTTGGCCACATTTAGCTATTTGGCGGCTCGCTCACTACTCTCGGCTCGAGACCAAAAGTCT	470.0
QY	4652	GCAGACGCGCAGAGCTCTGCTCTGCGGCCGCCAACCGAGCGAGCGGCGCAGAGAGGAG	471.1
Db	4701	CCAGACTGCTCGGCTCTGCGCGGCGGCGCAGAGTGAAGAGCGAGCGCATAGAGGAGAG	476.0
QY	4712	TGGGCAA 4718	
Db	4761	TGGCCAA 4767	
RESULT 15			
AAD00831/c			
ID	AAD00831 standard; DNA; 7215 BP.		
XX			
AC	AAD00831;		
XX			
DT	08-SEP-2000 (first entry)		
XX			
DE	Hybrid adeno-associated virus 2/4 helper plasmid sequence.		
XX			
KM	Hybrid virus plasmid; adeno-associated virus; AAV2; AAV4; gene therapy;		
KW	recombinant parvoviral vector; cellular tropism; rep gene; cap gene;		
KM	capsid; gene delivery; ss.		
XX			
OS	Chimeric - Adeno/associated virus serotype 2.		
XX	Chimeric - Adeno associated virus serotype 4.		
FT	Key		
FT	CDS		
FT	Location/Qualifiers		
FT	251..2183		
FT	/*tag= a		
FT	/*product= "Rep 68"		
FT	/*note= "The specification states that this region		
FT	encodes Rep 68 protein of AAV2. However, the positions		
FT	given in the specification for this CDS do not correspond		
FT	to start or stop codons"		
FT	251..2120		
FT	/*tag= b		
FT	/*product= "Rep 78"		
FT	/*note= "The specification states that this region		
FT	encodes Rep 78 protein of AAV2. However, the positions		
FT	given in the specification for this CDS do not correspond		
FT	to start or stop codons"		
FT	CDS		
FT	251..2120		
FT	/*tag= b		
FT	/*product= "Rep 78"		
FT	/*note= "The specification states that this region		
FT	encodes Rep 78 protein of AAV2. However, the positions		
FT	given in the specification for this CDS do not correspond		
FT	to start or stop codons"		

```

FT      CDS
FT      923..2120
FT      /*tag= C
FT      /product= "Rep 52"
FT      /note= "The specification states that this region
FT      encodes Rep 52 protein of AAV2. However, the positions
FT      given in the specification for this CDS do not correspond
FT      to start or stop codons"
FT      923..2183
FT      CDS
FT      923..2183
FT      /*tag= d
FT      /product= "Rep 40"
FT      /note= "The specification states that this region
FT      encodes Rep 40 protein of AAV2. However, the positions
FT      given in the specification for this CDS do not correspond
FT      to start or stop codons"
FT      2123..4341
FT      CDS
FT      2123..4341
FT      /*tag= e
FT      /label= "Cap protein region"
FT      /note= "The Cap protein region contains the coding
FT      regions for the proteins Vp1, Vp2 and Vp3 where the
FT      start positions for these CDS are 2123, 2547 and 2727
FT      respectively. However, the positions given in the
FT      specification for this CDS do not correspond to start
FT      or stop codons"
FT

```

```

XX MO200028004-A1.
FN 18-MAY-2000.
PD
XX
XX 10-NOV-1999; 99WO-US26505.
PF
XX 10-NOV-1998; 98US-0107840.
PR
XX 10-MAR-1999; 99US-0123651.
PR
XX (UTNC-) UNIV NORTH CAROLINA.
PA
XX
XX Rabinowitz JE, Samulski RJ, Xiao W;
PI
XX WPI: 2000-376523/32.
DR
XX
XX Recombinant parvoviral vectors with altered packaging, tropisms and
PT immunogenic properties, useful in gene therapy protocols -
XX
XX Example 28; Page 130-134; 153pp; English.
PS
XX
CC The patent discloses modified parvovirus vectors with advantageous
CC antigenic properties, packaging capabilities and cellular tropisms.
CC These vectors can be used in standard recombinant DNA protocols e.g. gene
CC therapy for delivering nucleic acids to cells. The present sequence is
CC a hybrid virus plasmid comprising adeno-associated virus serotype 2
CC (AAV2) rep genes and AAV4 capsid (cap) genes in a plusscript backbone.
CC The rep gene encodes four proteins responsible for replication,
CC Rep 68, Rep 78, Rep 52 and Rep 40 and the cap gene encodes three
CC structural proteins VP1, VP2 and VP3. The hybrid virus is used to
CC introduce a heterologous gene into a target cell.
XX
SQ Sequence 7215 BP; 1548 A; 1918 C; 1925 G; 1824 T; 0 other;
Query Match 51.7%; Score 2438.8; DB 21; Length 7215;
Best Local Similarity 75.0%; Pred. No. 0;
Matches 3273; Conservative 0; Mismatches 1012; Indels 77; Gaps 15
OY 205 GTCTGTATTAGCTGTCAAGTGTCTTTTGGACATTTTGGACACCAAGTGGCAATT 264
DB 5002 GTCTGTATTAGAGGTCAAGTGTG-TTTTGGACATTTTGGACACCAAGTGTGTCAG 4944
OY 265 TAGGGTATATATGGCCAGTAGGAGGAGAGAGATCTCCATTTTG--ACGGGAAATTTGAC 323
DB 4943 CTGGGTATTATTAAGCCGAGTAGAGACGAGAGGGTCTCCATTTTGAAGCCGGAGTTTGAAC 4884
OY 324 GAGAGAGAGCCATCGGGGCTTTTACGAGATCGATGATCAAGTGGCCGAGCACTTGACG 383
DB 4883 GCGAGCCGCGCATGCGGGGTTTTCACGAGATTGTGATTAAAGTCCCGACGACCTTGAAG 4824

```

QY	384	AACACCTGCGGGCAATTTCTGACTGTTTGTGAAGCTGGGTGGCCGAGAAAGAAATGGAGAC	443
Db	4823	GGCAATCTGCGCGCAATTTCTGACAGTTTGTGAACCTGGGTGCGCAGAAAGAAATGGAGAT	4764
QY	444	TGCCCCCGAATTCGACATGATCTGAATCTGATTTAGACAGGACACCCCTGACCTGTGGCG	503
Db	4763	TGCGCGCAGATTCTGACATGATCTGAATCTGATTTAGAGAGCACCCCTGACCTGTGGCG	4704
QY	504	AGAACTGAGAGGCCACTTCTGTGTCMAATGGCGCGCGGTAGTAAAGCCCTGGAGGCC	563
Db	4703	AGAACTGAGAGGCCACTTCTGTGTCMAATGGCGCGCGGTAGTAAAGCCCTGGAGGCC	4644
QY	564	TCTTCTTTGTCAGTTTCGAGAGGGCGAGTCTTACTTCCACTCCATATTTCTGTGGAGA	623
Db	4643	TTTTCTTTGTGCAATTTGAGAGGAGAGAGTACTTCCACATGACGTGCTGTGAGAA	4584
QY	624	CCACGGGGGTCAATTCATGTGTGTGGCGCTTCTTGATGATTAAGGACAAAGCTGG	683
Db	4583	CCACGGGGGTCAATTCATGTGTGTGGCGCTTCTTGATGATTAAGGACAAAGCTGG	4524
QY	684	TGCAGACCATTCACGGCGGGATTCGAGCGGACCTCTGCCAATCGTTTGGCGGATCAAGA	743
Db	4523	TTTCAAGGAATTTACCGCGGGATTCGAGCGGACCTTCTGCCAATCGTTTGGCGGATCAAGA	4464
QY	744	CGCGTAATGGCGCGGAGGGGGAAACAAGTGTGAGCAGAGTCTACATCCCACTACC	803
Db	4463	CCAGAAATGGCGCGGAGGGGGAAACAAGTGTGAGCAGAGTCTACATCCCACTACT	4404
QY	804	TGCTGCCCAAGCTACAGCCCGAGCTGACGTGGCGGTGACCTAACATGAGAGATTAATA	863
Db	4403	TGCTGCCCAAAACCAAGCTTGAAGCTCCAGTGGGCGGTGACCTAACATGAGATTTTAA	4344
QY	864	GCGCTGTTTGAACCTGGCGGAGCGCAACCGCTGTGTGGCGAGACCTGACCCACTCA	923
Db	4343	GCGCTGTTTGAATCTCAACGAGCGTAAACGGTGTGTGGCGAGACTGACCGACGCT	4284
QY	924	GCCAAACCCAGAGACAGAAACAAGAGATCTGAACCCCAATTTCTGACGCGCTGTATCC	983
Db	4283	CGCAACCCAGAGAGAGAAACAAGAGATCAAGAAATCCCAATTTCTGATCGCGCGGTATCA	4224
QY	984	GGTCAAAAACCTCCCGCGCTCATGAGAGTGTGGGTGGCTGTGTGACCGGGCATCA	1043
Db	4223	GATCAAAAACCTTCAACCGAGTACATGAGAGTGTGGGTGGCTGTGTGACCAAGGGATTA	4164
QY	1044	CCTCCGAGAGAGCTGATCCAGAGAGACAGGCGCTCGTACATCTCTTCAACGCGCTT	1103
Db	4163	CCTCCGAGAGAGAGTGAATCCAGAGAGACAGGCGCTCAATCATCTCTTCAATGGGCGCT	4104
QY	1104	CCAACTCGGGGTCCAGATCAAGGCGCGCTGTGACAAATGCGCGCAAGATCAATGGCGCTGA	1163
Db	4103	CCAACTCGGGGTCCAAATCAAGGCTGCTTGAACAAATGCGGAAAGATTAATGAGCTTGA	4044
QY	1164	CCAAATCGCGCGCGACTACTGTGAGGCCCGCTCGCGCGCGGACATTAACCAACC	1223
Db	4043	CTAAACCGCGCGCGACTACTGTGAGGGCGAGCAGCAGCGGTGAGGACATTTCCAGCAATC	3984
QY	1224	GCATTTACCGATCTGTGAGCTGAACGAGCTAGAACCTGCTCAGCGCGGCTCCGCTTTC	1283
Db	3983	GGATTTATTAATTTTGTGAACCTAAACGGGTAGATCCCAATATGCGGCTTCCGCTTTC	3924
QY	1284	TGCGCTGGGGCCAGAAAAGTTTCGGGAAGCGCAACACATCTGGCTGTTTGGCGCGCA	1343
Db	3923	TGGGATGGGGCCAGAAAAGTTTCGGCCAAAGAGAAACATCTGGCTGTTTGGCGCTGCA	3864
QY	1344	CCACGGGCAAGACCAACATTCGCGGAAGCCATGCGCCACGCGGTGCTTTCTACGCTGCG	1403
Db	3863	CTACCGGGAAGCAACCAATTCGCGGAGCGCAATGCCCACTGTGCTTTCTACGCGGTGCG	3804
QY	1404	TCAACTGAGCCAAATAGAACTTTCCTTCAATGATTTGGTGTGACAAAGATGGTGAATGGT	1463
Db	3803	TAAATCTGAGCCAAATAGAACTTTCCTTCAACGACTGTGTGTGACAAAGATGGTGAATGGT	3744
QY	1464	GGGAGAGAGGCAAGATGACGGCGCAAGGCTGTGGAATCGGCCAAGGCCATTTCTGGCGGCA	1523

Db	3743	GGAGAGAGGAGAGATGACCGCAAGGTCGTGAGATCGGCCAAGACCAATTTTCGAGGAA	3684
Qy	1524	GCNAGGTGCGCGTGGACCCAAAAGTGCATGTCGCCAGATGCACCCACCCCGTGA	1583
Db	3683	GCNAGGTGCGCGTGGACCCAAAATGCAATGTCCTCGGCCAGATGACCCGACTCCCGTGA	3622
Qy	1584	TGCTCACTCCCAACCAACATGTGCGCCGTGATTTGACGGGAACGACCAACTTTCGAC	1643
Db	3623	TGCTCACTCCCAACCAACATGTGCGCCGTGATTTGACGGGAACGACCAACTTTCGAC	3564
Qy	1644	ACCACGACCGGTGCGAGCACCGGATGTTCAATTGAACTCACCCGCGCTGTGAGCATG	1703
Db	3563	ACCACGACCGGTGCGAGCACCGGATGTTCAATTGAACTCACCCGCGCTGTGATCATG	3504
Qy	1704	ACTTTGCGAAGTGCACAAAGCAGAGATTCAAAGATTTCTCCGCTGGGCGCAGATCACG	1763
Db	3503	ACTTTGCGAAGTGCACAAAGCAGAGATTCAAAGATTTCTCCGCTGGGCGAAGATCACG	3444
Qy	1764	TGACCGGAGTGGCGCATGATGTTCTACGTCAAGAAAGGGTGGAGCCCAAAAGACCCGCC	1823
Db	3443	TGCTTGAAGTGGAGCATTAATTTCTACGTCAAAAGGGTGGAGCCCAAGAAAGACCCGCC	3384
Qy	1824	CCGATGACGCGGATPAAAAGCAGCCCAAGCGGGCCTGCTCACTGTCGCGATTCATCGA	1883
Db	3383	CCAGTGAACGAGATPATAAGTAGAGCCCAAGCGGGTGGCGAGTTCAGTTGCCAGCATCGA	3324
Qy	1884	CGTCAAGCGCGAAGAGACTCCGGTGGACTTTGCGACAGGTACCAAAACAATGTTCTC	1943
Db	3323	CGTCAAGCGCGAAGAGACTCCGGTGGACTTTGCGACAGGTACCAAAACAATGTTCTC	3267
Qy	1944	GTCAACGCGGCGATGCTTCAGATGCTGTTTCCCTGCAGACATGCGAGAAATGATCAGA	2003
Db	3266	GTCAACGCGGCGATGATCTGATGCTTTTCCCTGCCGCAATGCGAGAAATGATCAGA	3207
Qy	2004	ATTTCAATTTTGTCTTCAACGACGGGACGABAGACTGTTCAGATGCTTCCCGCGCTGT	2063
Db	3206	ATGTGAATTTTGTCTTCAACGACGGGCTGACGAGCTGTGCGAGTCTTCCC---CGGT	3150
Qy	2064	CAGATCTCAAC-----CGGTGCTGCAAAAGAGACGATTCGGAAAATCTGTGCCATTC	2117
Db	3149	CAGATCTCAACCCGTGTCTGTGCTGCAAAAAGCGGATTCAGAAATCTGTGCCATTC	3090
Qy	2118	ATCATCTGCTGGGCGGCGCTCCCGAATGCTTGTCTGGCCTGCGATCTGTCAACGTGG	2177
Db	3089	ATCATCATATGAGGAGGGCGGCCGACGAGGTGCTCTCGCCTGTGGAATCTGGCCATGTGG	3030
Qy	2178	ACCTGATGATCTGTGTTTCTGAGCAATPAAATGACTTAAACAGAGTATGTGCGCATGTGT	2237
Db	3029	ACTTGATGATCTGTGATGACATGGAACATPAAATGACTTAAACAGATATGATGACT---GACGT	2973
Qy	2238	TATCTTCAGATTGAGCTCGAGGAACTCTCTGAGGGCATTTGCGCATGTGTGGAATGG	2297
Db	2972	TACCTTCAGATTGAGCTCGAGGAACTCTCTGAGGGCATTTGCGCATGTGTGAGGACTGG	2913
Qy	2298	AAACCTGGAAGCCCGGAAGCCCAAGGCCAACAGAGAAAGACGAGACGACGCGCGGGGTCTG	2357
Db	2912	CAACTGGAAGCCCTTAAACCCAAAGGCAATCAACATCAAGGACAAAGCTCGGGGATCTT	2853
Qy	2358	GTGCTTCTGTGCTCAAGATACCTCGGACCTTCAACGGACTCGACAAAGGGGAGACCCGTC	2417
Db	2852	GTGCTTCTGTGCTCAAGATACCTCGGACCTTCAACGGACTCGACAAAGGGGAGACCCGTC	2793
Qy	2418	AAACGCGGAGGACGACGCGGCGCTCGAGACGACAAAGGCTTACACACGACGACTTAAAGCG	2477
Db	2792	AAACGAGGAGGACGCGGCGGCGCTCGAGACGACAAAGGCTTACACACGACGACTTAAAGCG	2733
Qy	2478	GATGCAATTCGTACTGCGGATTAATCAACGCGCGAGCTTCAGAGAGCTGTGCA	2537
Db	2732	GATGCAATTCGTACTGCGGATTAATCAACGCGCGAGCTTCAGAGAGCTGTGCA	2673
Qy	2538	GAAATACGTCCTTTGGGGGCAACCTCGGGGAGACGCTTTCAGGCTCAAGAGCGGCTT	2597

Db 2672 GGGGACATCGTTTGGGGCAACCTCGGAGAGACTCTTCCAGCCCAAAAAGAGGTT 2613
 QY 2598 CTCGAACCTCTCGGTCGTGTTGAGAAAGGCGCTAAGACGGCTCTGGAAGAAACGTCG 2657
 Db 2612 CTGGAACCTCTTGGTTGGTTGAGCAAGCGGTGAGACGGCTCTGGAAGAAAGACCG 2553
 QY 2658 GTAGAGAGTCGCGCAAGAGGAGAGACTCTCTCGGGCATCGGCAAGACGCGACAG 2717
 Db 2552 TTGATTGAATCCCCGAGAGCCGAGCTCTCCAGGGGTATCGGCAAAAAGCAAGCAG 2493
 QY 2718 CCGGCTAAAAAGACTCAATTTTGGTCAAGCTGGGCACTCAAGATGATCCCCGATCA 2777
 Db 2492 CCGGCTAAAAAGAGCTCGTTT-----CGAAACGAAACTGAGAGAGCGAGCGA 2442
 QY 2778 CAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCTGGGACCTTACTCAATGAGCTTCA 2837
 Db 2441 CCCCCGAGGAGATCAACTTCCGAGGCAATGCTGATG-----ACAGTGAATGCGTGCA 2388
 QY 2838 GGGGCTGGCGCACCAATGAGCAACAAATMAAGAGCGCGAGAGTGGGTAAATGCTCA 2897
 Db 2387 GCAGCTGGCGGAGCTGAGTGAAGGGCGGACAAAGTCCGATGAGTGGGTAAATGCTCG 2328
 QY 2898 GGAATTTGGATTTGCCATTTTCCATGAGTGGGCGAGAGTCACTACACACGACACCCG 2957
 Db 2327 GGTGATTTGGCATTTGCGATTCACCTGCTGAGGGCGACGTCACGACACGACGACCA 2268
 QY 2958 ACCGCGGCTTGGCCACCTCAATTAACCTCTCAAGCAAAATCTCAAGTCTTCAAG 3017
 Db 2267 ACCGCGGCTTGGCCACCTCAATTAACACACCTCTCAAGGAGCTCGA----- 2220
 QY 3018 GGGGCGGACGACCAACCACTCTTGGGTCAGACACCCCTGGGGGTATTTTGAATTC 3077
 Db 2219 GAGAGCTGAGTCGCAACCACTCAACGATTTCCACCCCTGGGGATCTTTGATTC 2160
 QY 3078 AACAGATTCACCTGCGCATTTTCCACGATGATGCGAGGAGTCACTCAACAAATTTG 3137
 Db 2159 AACCGCTTCCACCTGCGCATTTTCCACGATGATGCGAGGAGTCACTCAACAAATTTG 2100
 QY 3138 GGAATTCGCGGCGGAGAGACTCACTCAACCTCTCAACATTCGAGGAGGAGTCAAG 3197
 Db 2099 GGCATGCGACCCCAAGGCAATGCGGGTCAAAATCTTCAACATTCGAGGAGGAGTCAAG 2040
 QY 3198 AGCAATGATGCGGTCACAAACATGCTAATTAACCTTACGACGAGTTCAGTCTTCG 3257
 Db 2039 ACCTGCGAGCGCGGAGCAAGCGGTGATTAATTAACCTTACGACGAGTTCAGTCTTCG 1980
 QY 3258 GACTCGGAGTACCAAGCTTCCGTCGTCCTGCGCTCTGCGACCAAGGCTGCTCTCG 3317
 Db 1979 GACTCGTCTGACGAATGCGGTACGTCGATGATGATGCGGTCAAGGCGACGCTGCTCT 1920
 QY 3318 TTCCCGGCGGAGCGTTCATGATTCGCGCAATGCGGCTAC-----CTGAGGCTCAAC 3368
 Db 1919 TTTTCCCAACAGCTTTATGATGTCGCCAGTACGCGTACGTCGATGTCGCGGCAAC 1860
 QY 3369 AATGCGACGCAAGCGGTGAGAGTTCATCTTTTACTGCTGGAATTTTCCCTTCCAG 3428
 Db 1859 ACTTTCGAGCAACGATGACGAAGAAATGCTTCTTACGCTGAGATCTTCTTCCAG 1800
 QY 3429 ATGCTGAGAAAGGCGCAACATTTTACTGATCACTCTTTGAGAGAGTCTTTCAC 3488
 Db 1799 ATGCTGCGGAGTGGCAACATTTGAATTAAGTACGATTTTGAAGAGTGGCTTTCAC 1740
 QY 3489 AGGAGCTAGCGGAGCAAGCGGAGCGTGAAGCGGTGATGATCTCTCAATCGCAATAC 3548
 Db 1739 TCGATGTCGCGCAACAGCGAGCTGAGCGGTGATGATGATCTCTCAATCGCAATAC 1680
 QY 3549 CTGTATTAACCTGAA-----CAGAACTCAAAATCAAGTCCGAAAGTCCCAAAAGACTT 3604
 Db 1679 CTGTGGGAGTCTGCAATGACCAACACCGGAACCACTGATGTCGGGAGTCCGACAC 1620
 QY 3605 GCTGTTTAAAGCGGTGGTCTCCAGCTGAGATGTCGTTCAAGCCCAAAATCTGCTAG 3664
 Db 1619 AACCTTACCAAGTGGCGGCTACCACTTTTCAACCTTTAAAAAGAACTGGCTGCGG 1560

QY 3665 ACCCTGTATTCGCGAGAGCGGCTTCTTAAAAAAGAGACCAACAGCAATTT 3724
 Db 1559 GCCTTCAATCAAGACAGAGGCTTCTCAAAAGCTGCGCAATCAAACTACAGATTCCTCG 1500
 QY 3725 TACTGAGTCTGCTCTTCAAAATATAC-----CTCAATGGCGGTGAATC 3769
 Db 1499 CACCGGCTGAGACAGTCTCATCAAAATACGAGACGACACGACTCTGAGGAAATGAG 1440
 QY 3770 CATTCAACCCCGGACCTGATGAGCTTACCAAAAGAGCGCGGAGCTTCAAACTGATTTGA 3829
 Db 1439 TGCCTGACCCCGGAGCTTCAATGAGCAAGGCTGAGCTGCGGAGACGAGTTC---AG 1383
 QY 3830 CATGACGCGGTGTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAAACTGATTTGA 3889
 Db 1382 CAACAGCCAGCTCATTTTGGCGGCGCTTAAACGAAGAGCAACAGGCAACGCTACCCG 1323
 QY 3890 CAATGTCATGATTAACAGAGAGGAAATTAAGCCACTAACCTGTTGGCCACCGAAG 3949
 Db 1322 GACTGATCTTCACTCTGAGAGGAGGCTGGGAGCACCAACGCCACCGATAGGACAT 1263
 QY 3950 ATTGGGACCGTGGAGTCAATTTCCAGAGAGCAGACAGACCTCGGACCGGAGATGT 4009
 Db 1262 GTGGGCAACCTACCTGGCGGTGACCAAGAGCAACAGCACTGCGACCGTGAAGACT 1203
 QY 4010 GCATGCTATGGAGAGCACTTACCTGCGCATGGTGTGGCAAGATAGAGAGTGAACCTGAGAG 4069
 Db 1202 GAGAGCTTGGAGAGCGGTGCTGATGATGCTGCGCAAAACAGAGATTTACTACAGAG 1143
 QY 4070 TCCCATTTGGGCGCAAAATCTTCAACAGATGACACTTACCCGCTCTCTTATGG 4129
 Db 1142 TCCCATTTGGGCGCAAAATCTTCAACAGATGACACTTACCCGCTCTCTTATGG 1083
 QY 4130 CGGCTTTGAGCTCAAAACCGGCTCTGAGATCTTCAATCAAAACAGCGTGTCTG 4189
 Db 1082 TGGGTTTGGCTGAACACCGGCTCTCTCAAAATTTATCAAAACACCGGTACTG 1023
 QY 4190 GAATCTCGGCGGAGATTTTCAAGCTCAAAAGTTTGTCTTCAATCAATCAATCAATCA 4249
 Db 1022 GAATCTCGGAGAGCTTCAAGCTTCAATCAATCAATCAATCAATCAATCAATCAATCA 963
 QY 4250 AGGACAGTGAAGTGAATGATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGGA 4309
 Db 962 TGGCGAGTGTGCGGTGAGATGATGAGAGATCCAGAAAGAGCGGTGCCAAAGCTGGAA 903
 QY 4310 TCCGAAAGTGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 4369
 Db 902 CCGGAGGTCAGATTTTACCTCAACAGAGAGCAAACTCTCTGTGGGCTCCGA 843
 QY 4370 CAACATGAGCTTATATGAGGCTGCGCCATTTGACACCGGTTACTTACCGGCTCCCT 4429
 Db 842 TGGGCGTGGAAATACAGTGAAGCTGAGGCTATGCGGTACCGGCTACTCACCAACACT 783
 QY 4430 GATATTAAGTGTATCAATCAATCAACCGGTTGATTTGATTTGATTTGATTTGATTTG 4488
 Db 782 GATATTAAGTGTATCAATCAATCAACCGGTTGATTTGATTTGATTTGATTTGATTTG 723
 QY 4489 TCCCTTCTTATCTTATC-GGTTACCATGTTATGCTTACCA 4529
 Db 722 TCCCTTCTTATCTTATC-GGTTACCATGTTATGCTTACCA 681

Search completed: July 20, 2003, 09:39:47
 Job time : 1083.76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 13:19:56 ; Search time 1013.56 Seconds
(without alignments)
9603.029 Million cell updates/sec

Title: US-09-807-802a-1

Perfect score: 4718
Sequence: 1 ttggccaccctctctctg.....cgcagagagagtgagtgcaaa 4718

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3107.8	65.9	4679	10	US-09-804-898-1
2	3107.8	65.9	4679	10	US-09-945-681-10
3	3107.8	65.9	4679	15	US-10-038-972A-12
4	3082.4	65.3	4680	15	US-10-077-294-1
5	3082.4	65.3	4680	15	US-10-163-886-1
6	3082.4	65.3	4680	15	US-10-263-127-1
7	3055.8	64.8	4675	11	US-09-782-378A-1
8	3055.8	64.8	4675	11	US-09-782-378A-2
9	3055.8	64.8	4675	15	US-10-240-198-1
10	2941.8	62.4	8179	15	US-10-205-942-5
11	2730	57.9	8151	15	US-10-205-942-2
12	2446.2	51.8	7214	15	US-10-205-942-1
13	1872	39.7	1872	11	US-09-792-630-12
14	1872	39.7	1872	12	US-09-953-351-12
15	1872	39.7	1872	15	US-10-080-376-12
16	1872	39.7	1872	15	US-10-082-671-18

17	1872	39.7	1872	15	US-10-097-100-12	Sequence 12, Appl
18	1872	39.7	1872	15	US-10-023-208-12	Sequence 12, Appl
19	1814.4	38.5	1872	11	US-09-792-630-14	Sequence 14, Appl
20	1814.4	38.5	1872	12	US-09-953-351-14	Sequence 14, Appl
21	1814.4	38.5	1872	15	US-10-080-376-14	Sequence 14, Appl
22	1814.4	38.5	1872	15	US-10-082-671-20	Sequence 20, Appl
23	1814.4	38.5	1872	15	US-10-097-100-14	Sequence 14, Appl
24	1814.4	38.5	1872	15	US-10-023-208-14	Sequence 14, Appl
25	1451.2	30.8	1875	11	US-09-792-630-8	Sequence 8, Appl
26	1451.2	30.8	1875	12	US-09-953-351-8	Sequence 8, Appl
27	1451.2	30.8	1875	15	US-10-080-376-8	Sequence 8, Appl
28	1451.2	30.8	1875	15	US-10-082-671-14	Sequence 14, Appl
29	1451.2	30.8	1875	15	US-10-097-100-8	Sequence 8, Appl
30	1451.2	30.8	1875	15	US-10-023-208-8	Sequence 8, Appl
31	1446.4	30.7	1875	11	US-09-792-630-10	Sequence 10, Appl
32	1446.4	30.7	1875	12	US-09-953-351-10	Sequence 10, Appl
33	1446.4	30.7	1875	15	US-10-080-376-10	Sequence 10, Appl
34	1446.4	30.7	1875	15	US-10-082-671-16	Sequence 16, Appl
35	1446.4	30.7	1875	15	US-10-097-100-10	Sequence 10, Appl
36	1446.4	30.7	1875	15	US-10-023-208-10	Sequence 10, Appl
37	1444.8	30.6	1872	11	US-09-792-630-6	Sequence 6, Appl
38	1444.8	30.6	1872	12	US-09-953-351-6	Sequence 6, Appl
39	1444.8	30.6	1872	15	US-10-080-376-6	Sequence 6, Appl
40	1444.8	30.6	1872	15	US-10-082-671-12	Sequence 12, Appl
41	1444.8	30.6	1872	15	US-10-097-100-6	Sequence 6, Appl
42	1444.8	30.6	1872	15	US-10-023-208-6	Sequence 6, Appl
43	1405	29.8	1932	14	US-10-022-339-587	Sequence 587, App
44	1405	29.8	1932	14	US-10-022-339-605	Sequence 605, App
45	1405	29.8	1932	14	US-10-022-339-637	Sequence 637, App

ALIGNMENTS

RESULT 1
US-09-804-898-1
; Sequence 1, Application US/09804898
; Patent No. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: DURING, MATTHEW
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804, 898
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189, 110
; SOFTWARE: PatentIn version 3.0
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 1
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-804-898-1

Query Match	65.9%	Score 3107.8;	DB 10;	Length 4679;
Best Local Similarity	80.2%	Pred. No. 0;		
Matches 3790;	Conservative	0;	Mismatches 882;	Indels 53;
			Gaps	10;
QY	1	TTGGCCACTCCCTCTGCGCGCTGCGTGGTGCGGAGTCCG	60	
DB	1	TTGGCCACTCCCTCTGCGCGCTGCGTGGTGCGGAGTCCG	60	
QY	61	AGACGGCAAGCTCTGCTGCGCGCCACCGAGCGAGCGAGGAGTG	120	
DB	61	CGACGCCCGCGGCTTTCGCCCGCGCGCTGAGGAGCGAGCGAGGAGTG	120	
QY	121	GGCAATCATCATCTAGGGGATATGCGGAAGGCGCTCCAGCGTCCGAGCGCTA	180	
DB	121	GCTCACTCATCATCACTAGGGGATATGCGGAAGGCGCTCCAGCGTCCGAGCGCTA	180	
QY	181	CGTAATTCAGTCATGAGG---GAGTGCTCTGATATTAAGTGTCAAGTGTCTTTC	237	

Db 164 CGTGAATTAGCTCATAGGGTTAGGGAGGTCCTGATTATTAAGGTCAACGTAAGTG-TTTTGC 222
Qy 238 GACATTTTGGCAACACAGTGGCAATTTAGGGTATATATGCGCAATGAGGACAGAT 297
Db 223 GACATTTTGGCAACACAGTGGTCAACGTTGGTATTTAAGCCCAAGTGAAGACCCAGAGGT 282
Qy 298 CTCATTTTGG-ACGCGCAATTTGAACGAGCAGCAGCATGCGGGGCTTCAAGAGATCG 356
Db 283 CTCATTTTGAAGCGGAGGTTTGAACGCGACGCGCCATGCGGGGTTTACAGAAATG 342
Qy 357 TGATCAAGTGCAGGAGCCTGAGCAGCACTGCGGGCAATTTGACTGTTTGA 416
Db 343 TGATTAAGTCCAGGACCTTACAGCAGCATCTGCGGCAATTTGACAGCTTTTGA 402
Qy 417 GCTGGGTGGCGGAGAAATGGAGCTGCCCCGGATTTGACATGATTTGAATCTGA 476
Db 403 ACTGGGTGGCGGAGAAATGGAGTGGCGCAGATTTGACATGATTTGAATCTGA 462
Qy 477 TTGAGCAGGCAACCCCTGACCGTGGCGGAGATGACGCGCACTTCTGGTCCAATGCG 536
Db 463 TTGAGCAGGCAACCCCTGACCGTGGCGGAGATGACGCGCACTTCTGACGAAATGCG 522
Qy 537 GCGCGTGAATAGAGCCCGGAGGCGCTTCTTTTGTTCAGTTTGAAGAGGCGAGTCT 596
Db 523 GCGGTGAATAGAGCCCGGAGGCGCTTCTTTTGTTCAGTTTGAAGAGGCGAGTCT 582
Qy 597 ACTTCCACCTCCATATTTGTGTGAGACACAGGGGGTCAATTCATGCTGTGGCGCT 656
Db 583 ACTTCCACATGCAAGTCTGTGAAACACCGGGGGTCAATTCATGCTTTTGGAGCTT 642
Qy 657 TCCGAGTCAAGTATGAGGACAAAGTGTGAGACCATCTACCGGGGATCGAGCCGACC 716
Db 643 TCCGAGTCAAGTATGAGGAAACCTGATTCAGAAATTTACCGGGGATCGAGCCGACTT 702
Qy 717 TGCCCAACTGGTTCGCGGTGACCAAGACGCGTAATGCGCGCGGAGGAGAAACAAGTGG 776
Db 703 TGCCCAACTGGTTCGCGGTGACCAAGAACAGAAATGCGCGCGGAGGAGAAACAAGTGG 762
Qy 777 TGGAACAGTGTATATCCCACTACCTCTGCCCCAGACTACACCGGAGTGTGAGTGG 836
Db 763 TGGAATGAGTGTATATCCCACTACCTCTGCCCCAGAAACCAACCTGAGTGTGAGTGG 822
Qy 837 CGTGAATCAATGAGGAGATATATAGCGCTGTTTGAACCTGCGCGAGGCAAAAGCGC 896
Db 823 CGTGAATCAATGAGGAGATATATAGCGCTGTTTGAATCTCAACGAGGATTAAGCT 882
Qy 897 TCGTGGCGGAGCCTGACCAAGTCAAGCCAGAGCCAGAGCAGAAACAAGAAATCTGA 956
Db 883 TCGTGGCGGAGCCTGACCAAGTCAAGCTGTGCGAGACGCGAGGAGCAAAACAAGAAATCTGA 942
Qy 957 ACCCCATTTTGAAGCGCGCTGTACCTCGGTCAAAAACCTCGCGCGCTACATGAGAGTGG 1016
Db 943 ATCCCAATTTTGAAGCGCGGTGTACAGATCAAAAACCTTCAGCAGGTACATGAGAGTGG 1002
Qy 1017 TCGGGTGGTGTGAGACCGGGGATCACTCCAGAAAGAGTGTGATCAAGAGAGCAGG 1076
Db 1003 TCGGGTGGTGTGAGACAGGGGATTACTTCGAGAAAGAGTGTGATCAAGAGAGCAGG 1062
Qy 1077 CTTGTACATCTCTTCAACGCGCTTCAACCTCGCGGTCCAGATCAAGCGCGCTGTGG 1136
Db 1063 CTTGTACATCTCTTCAACGCGCTTCAACCTCGCGGTCCCAATCAAGGCTCTCTTGG 1122
Qy 1137 ACATATGCGGCAATATGAGCGGTGACCAAAATCGCGCGCGGATTAATCTGTAGCCCG 1196
Db 1123 ACATATGCGGCAATATGAGCGGTGACCAAAACCGCGCGGATTAATCTGTAGCCCG 1182
Qy 1197 CTCGCGCGGAGCATTAACCAACGCACTCAACGATCTGAGAGTGAACGCTAGC 1256
Db 1183 AGCCCGGTGAGGACATTTCCAGCAATCGGATTTAATAATTTTGAACATAACGCGTACG 1242
Qy 1257 AACCTGCTTACGCGGCTCGGTCTTTCTCGGCTGGGCGCGAGAAAGGTTGGGAGCGCA 1316
Db 1243 ATCCCAATATGCGGCTTCTGCTTCTGAGATGGGCGACGAAAGATTTGCGCAAGAGGA 1302

Qy 1317 ACACCATCTGGCTGTTTGGCGCGGCAACGAGGCAAGACCAATCTGGGAGGCAATCG 1376
Db 1303 ACACCATCTGGCTGTTTGGCGCTGCAACTCCGGAAGACCAATCTGGGAGGCAATCG 1362
Qy 1377 CCCAGCGGCTGCTTCAAGGCTGTGCTCACTGACCAATGAGAACTTTTCTTCAATG 1436
Db 1363 CCCAGCTGTGCTTCTACGGGTGGGTAACTGAGCAATGAGAACTTTTCTTCAACG 1422
Qy 1437 ATTGCGTCAAGATGTGTATCTGTGGGAGGAGGCAAGATGACGCGCAAGTGTGG 1496
Db 1423 ACTGTGTCAAGATGTGTATCTGTGGGAGGAGGAGGAGATGACCGCAAGTGTGG 1482
Qy 1497 AGTCCGCAAGGCAATTTCTGGCGGAGCAAGGTTGGCGCGGAGCAAAATGCAAGCT 1556
Db 1483 AGTCCGCAAGGCAATTTCTGGAGAGAGCAAGGTTGGCGGTGACAGAAATGCAAGTCT 1542
Qy 1557 CCGCCAGATGACCCCAACCCCGTATGCTCACTCAACCAACATGTGGCGCGTGA 1616
Db 1543 CCGCCAGATGACCCGACTCCGTAATGCTCACTCAACCAACATGTGGCGCGTGA 1602
Qy 1617 TTGACGGGAAACAGACCACTTCAAGACCAAGCGCGTTGACAGACCGGATTTCAAT 1676
Db 1603 TTGACGGGAACTCAAGACCTTCAAGACCAAGCGCGTTGCAAGACCGGATTTCAAT 1662
Qy 1677 TTGAATCAACCGCGCTGAGAGCATGCTTTGGCAAGGTCAGAAAGAGGAGTCAAG 1736
Db 1737 AGTTCTTCCGCTGGGCGCAGGATCAAGTACAGTGAACGAGGTGGCGCATGATTTACGTCAAG 1796
Qy 1723 ACTTTTTCGGGTGGGCAAGGATCAAGTGTGAGGTGAGCATGATTTTCACTGCAAAA 1782
Db 1797 AGGTGAGGCAACAAAGACCCCGCGGATGACGCGGATTAAGCGAGCCCAAGCGGG 1856
Qy 1783 AGGTGAGGCAACAAAGACCCCGCGGATGACGCGGATTAAGGTGAGGAGCCCAAGCGGG 1842
Db 1857 CTTGCCCTCAAGTGGGAGTCCATGACGCTCAAGCGGAGGAGGCTCCGGTGAACCTTGG 1916
Qy 1843 TCGCGAGTCAAGTGGGAGGAGTCCATGACGCTCAAGCGGAGGAGGCTCCGGTGAACCTTGG 1899
Db 1917 CCGACAGTACCAAAAACAATGTTCTGTGACCGCGGATGCTTCAAGATGCTTTCCT 1976
Qy 1900 CAGACAGTACCAAAAACAATGTTCTGTGACCGTGGGATGATGATGCTTTCCT 1959
Db 1977 GCAAGATGCGGAGAAATGAAATCAAGATTTCAACTTTGCTTCAAGCGGAGCAGAG 2036
Qy 1960 GCAAGATGCGGAGAAATGAAATCAAGATTTCAACTTTGCTTCAAGCGGAGCAGAG 2019
Db 2037 ACTGTTCAGAGTCTTCCCGCGGTGTCAGAAATCTCAACCGGT---GTCAGAAAGAGA 2093
Qy 2020 ACTGTTCAGAGTCTTCCCGCGGTGTCAGAAATCTCAACCGGT---GTCAGAAAGAGA 2076
Db 2094 CGTATCGAAACTCTGTGCCATTCATATCTGTGAGGCGGAGTCCCGAATGCTTGTCT 2153
Qy 2077 CGTATCGAAACTCTGTGCCATTCATATCTGTGAGGAGGAGGAGGAGGAGGAGGAGG 2133
Db 2154 CCGCTGTGATCTGTGTCAAGTGAACCTGATGATCTGTGTTTGTGAGCAATAATGACTT 2213
Qy 2134 CTTGCTGCGATCTGTGTCAAGTGAATTTGATTAATCTGCACTTTTGAACAATAATGATTT 2193
Db 2214 AAACAGAGTATGCTGCGGATGTTATCTTCAAGATGAGTGTGAGGCAACCTCTGAG 2273
Qy 2194 AAATCAGATGAGTGTGCGGATGTTATCTTCAAGATGAGTGTGAGGCAACCTCTGAG 2253
Db 2274 GGCATTCGAGATGAGTGGGATTTGAAACCTGAGCCCGGAGGCAAGGCAACAGCA 2333
Qy 2254 GGAATTAAGACAGTGTGAGGAGGCTCAACCTGAGCCCAACCAACCAAGGCGGAGGCGG 2313
Db 2334 AAGCAGAGCAGCGGCGGAGTGTGTGCTTCTGTGCTCAAGATCTCTGAGACCTTCAAC 2393
Qy 2314 CATTAAGAGCAGCAGGAGGCTTGTGTCTTCTGTGAGTGAACAATGATCTCTGAGACCTTCAAC 2373

QY 2394 GGAATCGACAGGGGAGCCGCTCAACGCGGAGCCGAGCCGCTCGAGCAGACAG 2453
DB 2374 GGAATCGACAGGGGAGCCGCTCAACGCGGAGCCGAGCCGCTCGAGCAGACAG 2433
QY 2454 GCGTACGACGAGCTCAAGCGGAGTGAACATCTGAGCTGCGATTAACAAGCCGAC 2513
DB 2434 GCGTACGACGAGCTCAAGCGGAGTGAACATCTGAGCTGCGATTAACAAGCCGAC 2493
QY 2514 GCGGAGTTTCAAGAGCGCTGCAAGAGATACGCTTTTGGGGGCAACCTCGGGGAGCA 2573
DB 2494 GCGGAGTTTCAAGAGCGCTTAAAGAGATACGCTTTTGGGGGCAACCTCGGGGAGCA 2553
QY 2574 GCTCTCAGGCGCAAGAGCGGCTTCTGAACTCTCGGCTGCTGAGAGAGCGCTAAG 2633
DB 2554 GTCTTCAGGCGCAAGAGAGGCTTCTGAACTCTCGGCTGCTGAGAGAGCGCTTAAAG 2613
QY 2634 ACCGCTCTGGAAGAAAGCTCCGGTGAAGAGTGCACAAAGAGCAACTCTCTCTCG 2693
DB 2614 ACCGCTCTGGAAGAAAGAGCGGCTGAGACACTCTCTGAGGCGCAACTCTCTCTCG 2673
QY 2694 GCGATCGGCAAGACAGGCGGAGCCGCTAAAGAGACTCAATTTTGGTCAAGACTGCG 2753
DB 2674 GGAACCGGAAGGCGGCGGAGCGCTGCAAGAAAGATTAATTTTGGTCAAGACTGGA 2733
QY 2754 GACTCAGAGTCAAGTCCCGGATCCCACTCTCGGAGAACTCCGAGCAACCCCGCTGCT 2813
DB 2734 GAGCGAGACTCAAGTCAAGTCCCGGATCCCGGAGAACTCCGAGCAACCCCGCTGCT 2793
QY 2814 GTGGGAGCTTCAATGAGCTTCAAGCGGAGTGCAGCAATGAGCAATTAAGAAAGG 2873
DB 2794 GTGGGAGCTTCAATGAGCTTCAAGCGGAGTGCAGCAATGAGCAATTAAGAAAGG 2853
QY 2874 GCGGAGGAGTGGGATATGCTTCAAGAAATTTGGATTTGCAATGAGTGGGAGGAC 2933
DB 2854 GCGGAGGAGTGGGATATGCTTCAAGAAATTTGGATTTGCAATGAGTGGGAGGAC 2913
QY 2934 AGAGTCAATCAACAGCAGCAGCCGCACTGGGCGCTTGGCCACTTAATTAACAGCTTAC 2993
DB 2914 AGAGTCAATCAACAGCAGCAGCCGCACTGGGCGCTTGGCCACTTAATTAACAGCTTAC 2973
QY 2994 AAGCAATCTCAAGTCTTCAAGCGGAGGAGCAAGCAACCACTTTCGCTACAGC 3053
DB 2974 AAGCAATCTTCAAGCAGCTCAAGCGGAGGAGCAAGCAACCACTTTCGCTACAGC 3030
QY 3054 ACCCGCTGGGGGATTTTGAATTTCAACAGATTCAGCTTTCACAGCTGAGTGG 3113
DB 3031 ACCCGCTGGGGGATTTTGAATTTCAACAGATTCAGCTTTCACAGCTGAGTGG 3090
QY 3114 CAGGAGCTCAATCAATGAGGAGATTCGCGGCGCAAGAGACTCAACTTCAACTCTTC 3173
DB 3091 CAGGAGCTCAATCAATGAGGAGATTCGCGGCGCAAGAGACTCAACTTCAACTCTTC 3150
QY 3174 AAGATCAATCAAGAGAGTCAAGCAAGATGAGTGCATCAACATCGTAAATTAACCTT 3233
DB 3151 AAGATCAATCAAGAGAGTCAAGCAAGATGAGTGCATCAACATCGTAAATTAACCTT 3210
QY 3234 ACCGAGCAGGTTCAAGTCTTCTGAGACTCGAGATACAGCTTTCGATGCTGAGCTCT 3293
DB 3211 ACCGAGCAGGTTCAAGTCTTCTGAGACTCGAGATACAGCTTTCGATGCTGAGCTCT 3270
QY 3294 GCGGAGCAGGAGCTTCCGCTCGCTTCCGCGGAGAGTTCATGATTCGCAATTAAGGCG 3353
DB 3271 GCGGAGCAGGAGCTTCCGCTCGCTTCCGCGGAGAGTTCATGATTCGCAATTAAGGCG 3330
QY 3354 TACCTGAGCTCAACATGAGCAGCAAGCCGAGGAGCTTCACTCTTTTACGCTGAGAA 3413
DB 3331 TACCTGAGCTCAACATGAGCAGCAAGCCGAGGAGCTTCACTCTTTTACGCTGAGAA 3390
QY 3414 TATTTCTCTTCTCAAGTCTGAGAAAGCGGCAACATTTTACGCTTCACTCTTGGAG 3473
DB 3391 TATTTCTCTTCTCAAGTCTGAGAAAGCGGCAACATTTTACGCTTCACTCTTGGAG 3450
QY 3474 GAGGCTCTTTCACAGAGCTAGCGGCAAGCGAGCTGAGCGGCTGATGAATCTT 3533

DB 3451 GAGCTTCTTTCACAGAGCTTACGCTCAAGCAGAAAGCTGAGCCGCTCATGAATCTT 3510
QY 3534 CTATATGACCAATATCTGTATTAATCTGAAACAGAACTCAAAATATGATGCGAGTCCAA 3593
DB 3511 CTATATGACCAATATCTGTATTAATCTGAAACAGAACTCAAAATATGATGCGAGTCCAA 3570
QY 3594 AACAGGATCTGTGTATGAGCGGAGTCTCCAGCTGAGATGCTGTAGAGCCCAAAAC 3653
DB 3571 CAGTCAAGGCTTCAATGATTTCTAGGCGGAGGAGTGAATTCGGGACCAAGTCTTAGAAC 3630
QY 3654 TGGCTACCTGAGCCCTGTATGAGCAGAGCGGCTTCTTAAACAAACAAACAGACAAAC 3713
DB 3631 TGGCTTCTGAGCCCTGTATACGCGCAGAGGAGTATCAAAAGCATCTGCGATTAACAC 3690
QY 3714 AACAGCAATTTTACTGAGACTGCTTCAAAATATTAACCTCAATGAGGAGTGAATCAATC 3773
DB 3691 AACAGCAATTTACTGAGACTGAGCTTCAAAATATTAACCTCAATGAGGAGTGAATCAATC 3750
QY 3774 ATCAACCTTGGGACATGCTATGAGCTTCAACAAAGAGAGAAAGCAAGTCTTTCATG 3833
DB 3751 GTGAAATCCGGCCCGGCAATGAGCAAGCAAGAGAGATGAAGAAAGTCTTCTCAG 3810
QY 3834 AGCGGTGCTATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACTGATTTGACAAAT 3893
DB 3811 AGCGGTGCTTCTATCTTTGGAAAGAGAGCTCAGAGAAACAAATGAGCAATTTGAAAG 3870
QY 3894 GTGATATTAACAGAGAGAGAAATTAAGCACTAACCTTGTGCGACCGAGATTT 3953
DB 3871 GTGATATTAACAGAGAGAGAAATTAAGCACTAACCTTGTGCGACCGAGAGTAT 3930
QY 3954 GGGACGCTGAGCTCAATTTTCCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4013
DB 3931 GGTTCGTATCTACCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3990
QY 4014 GCTATGAGAGCAATTAAGTGGAGAGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4073
DB 3991 ACAGAGGAGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4050
QY 4074 ATTTGGGCAAAATTTCTTCAACAGATGAGCACTTACCCGCTCTCTTATGAGGAGG 4133
DB 4051 ATTTGGGCAAAATTTCTTCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4110
QY 4134 TTTGAGCTGAAGAGCGGCTCTGAGATCTTCAATTAAGAGAGAGAGAGAGAGAGAGAG 4193
DB 4111 TTTGAGCTGAAGAGCGGCTCTTCAAGATTTCTATTAAGAGAGAGAGAGAGAGAGAG 4170
QY 4194 CTTCCGCGGAGTCTTCAAGTCAAGAGTGTCTTCAATCAACCAATATCTTCAAGAG 4253
DB 4171 CTTCCAGCACTTCAAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4230
QY 4254 CAGTGAAGTGAAGATTAATGAGAGGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAG 4313
DB 4231 CAGTGAAGTGAAGATTAATGAGAGGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAG 4290
QY 4314 GAGTGAAGTGAAGATTAATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4373
DB 4291 GAGTGAAGTGAAGATTAATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4350
QY 4374 AATGAGCTTATCTGAGGCTGCGCCATTTGGAGCCGTTTACCTTACCTTCCGCTTAA 4433
DB 4351 AATGAGCTTATCTGAGGCTGCGCCATTTGGAGCCGTTTACCTTACCTTCCGCTTAA 4410
QY 4434 TTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4493
DB 4411 TTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4470
QY 4494 CTATATGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4553
DB 4471 TCTTCTTATCTAGTGTTCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4530
QY 4554 ATTAAGAGCTTACGATGAGGAGTACCTTACGATGAGGAGTGAAGTGAAGTGAAGTGAAG 4613

Db	451	TACAAGAG-----ACCCCTAGTAGAGGTGGCCATCCCTCTCTCG	4574
Qy	4614	CGCTCGCTCGCTCGGTGGGGCCCTGGAGAACAAAGTCCGCAACGGCAGAGTCTGCT	4673
Db	4575	CGCTCGCTCGCTCACTAGAGCCGGGCGAACAAAGTCCGCCAGCCCGGGCTTGGCCG	4634
Qy	4674	GCGGCGCCCAACCGAGCGACGAGCGCGCAAGAGGAGGTGGACAA	4718
Db	4635	GCGGCGCTCAGTAGAGCGAGCCCGCAAGAGGAGGTGGCAAA	4679

```

RESULT 2
US-09-945-681-10
: Sequence 10, Application US/09945681
: Patent No. US20020064878A1
: GENERAL INFORMATION:
: APPLICANT: UNIVERSITE DE NANTES
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
: FILE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION
: FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES
: CURRENT FILING DATE: 2001-09-05
: PRIOR APPLICATION NUMBER: US/09/945,681
: PRIOR FILING DATE: 2000-03-05
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10
: LENGTH: 4679
: TYPE: DNA
: ORGANISM: adeno-associated virus 2
US-09-945-681-10.

```

Query Match	65.9%	Score	3107.8	DB	10	Length	4675
Best Local Similarity	80.2%	Pred. No.	0				
Matches 3790; Conservative	0	Mismatches	882	Indels	53	Gaps	10

OY	1	TTGGCCATCTCCCTCTCTGCGAGCTGAGTCTGCTCGGTGGAGGACTGTGGAGCAAAAGGTCGGC	60
OY	1	TTGGCCATCTCCCTCTCTGCGAGCTGAGTCTGCTCGGTGGAGGACTGTGGAGCAAAAGGTCGGC	60
Db	1	TTGGCCATCTCCCTCTCTGCGAGCTGAGTCTGCTCGGTGGAGGACTGTGGAGCAAAAGGTCGGC	60
OY	61	AGACGGCAGAGCTGTCTCTGCGCGGCCCCACCGAGCGAGCGAGCGCGAGAGGAGATG	120
Db	61	CGAGCGCCGGGCTTTGGCCCGGGCGGCTCACTGAGCGAGCGAGCGCGAGAGGAGATG	120
OY	121	GCGCACTCCATCTACTAGGGGGTATATCGGAAAGCGCTCCACGCTGCGCGCTCAGCGCTGA	180
Db	121	GCGCACTCCATCTACTAGGGGGTATATCGGAAAGCGCTCCACGCTGCGCGCTCAGCGCTGA	180
OY	181	CGTAAATTACACGCANAGG---GAGTGGTCCGTATTTAGTCTACCGTGAAGTCTTTGC	237
Db	164	CGTGAATTACGCATAGGGTTAGGGAGGTCCTGTATTAAGTCACTGAGATG-TTTTGC	222
OY	238	GACATTTTGCAGACACACAGTGGCCATTATAGGGTATATATGGCCAGTAGCGACGACGAT	297
Db	223	GACATTTTGCAGACACACAGTGGTATCGTGGTATTTAAGCCGAGTAGACACGAGGTT	282
OY	298	CTCCATTTTG-AACGCGAAATTGAAAGAGCAGCAGCAGCATGCGGGCTTTTACGAGATCG	356
Db	283	CTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCGCATGACGGGGTTTACGAGATTTG	342
OY	357	TGATCAAGSTGCCAGAGCACTTGACAGAGCAGCTGCGGGCATTTTCTAGCTCTGTTTGA	416
Db	343	TGATTTAAGTCCCGAGCACCTTTCAGAGCATCTGCCCCGGCATTTTCTACAGCTTTTGA	402
OY	417	GCTGGGTGGCCGAGAAAGGATGGGAGCTGCCGCCCGGATTTCTGACATGGAATCGAATCGA	476
Db	403	ACTGGGTGGCCGAGAAAGGATGGGAGTTCGCCGAGATTTCTACATGGAATCTGAATTCGA	462
OY	477	TTGAGCAGGACCCCTTGACCGTGGCCGAGAAAGCTCAGCGCACTCTTGGTCTCCATGGC	536
Db	463	TTGAGCAGGACCCCTTGACCGTGGCCGAGAAAGCTCAGCGCACTTTTCTGAGGAATGGC	522
OY	537	GCCGGTAGTAAAGCCCCGAGAGCCCTCTTTTGTTCAGTTCGAAAGGAGCAGTCTT	596

Db	523	GCCTGTGAGTAAGCCCCCGAAGCCCTTTTCTTTGTGCAATTGAGAAAGGAGAGACT	582
Qy	597	ACTTTCACCTTCATATATCTGTGTGAGACCAACGGGGTCAATTCATGTGTCTGCGCCGT	656
Db	583	ACTTCACATGTCAGTGTCTCGTGGAAACCAACGGGGGTGAATTCATGTGTTTGGGACGTT	642
Qy	657	TCCTGAGTCAGATTAGGGACAAAGTGGTGCAGACCACTACCCGGGGATGTGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTCGGGAAAAATGATTAGAGAAATTTACCGGGGATGAGCCGACTT	702
Qy	717	TGCCCAATCGGTTTGGCGGTGACCAAGACGGGTAATGCGCGCGGAGGGGGAGCAAGGCTG	776
Db	703	TGCCAAATGTGTTTGGCGGTCACAAAAGACAGAAATGCGCGCGGAGGGGGAGCAAGGCTG	762
Qy	777	TGACGAGGTGTACATCCCACTACCTCTGTGCCAAGACTAGCCCGAGCTGTGAGTGG	836
Db	763	TGATGTAGTGTATCATCCCAATTACTTGTCTCCCAAAACCAAGCTGACTCGAGTGG	822
Qy	837	CGTGTACTAATGTAGAGAGATATTAAGCGCTGTTTGAACTGTGCGGACCGCAAAAGCG	896
Db	823	CGTGACTTAATGTGAACAGATATTTAAGCGCTGTTTGAAATCTCACCGAGCCGTAAAGGT	882
Qy	897	TCGTGGCGAGCACTTACCCACGCTGACCGACAGACCCAGAGACAGAACAGAGAACTGA	956
Db	883	TGTGTGGCGAGCATCTTAACGCACTGTTCGAGACGCGAGGAGCAGAACAAAGAGATAGA	942
Qy	957	ACCACAAATTCGACGCGCCTGTCAATCCGATCAAAAACTTCGCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTCGATGCGCCGGTGTATGATATCAAAAACTTCAGCAGAGTATCATGAGCTGG	1002
Qy	1017	TCGGTGTGCTGTGTGACCCGGGGCATACCTTCGAGAAAGCAGTGTATCAGAGAGACAGG	1076
Db	1003	TCGGTGTGCTGTGTGACAAAGGGATTTACTTCGAGAAAGCAGTGTATCAGAGAGACAGG	1062
Qy	1077	CCTGTACATCTCCTTCAACGCGCTTCAACTCGCGGTGCCAGATCAAGCGCTGTGG	1136
Db	1063	CCTATACATCTCCTTCAATGCGGCTTCCAACTGCGGTTCCAAATCAAGGCTGTGG	1122
Qy	1137	ACAATGCGCGCAAGATCATGCGCTGTGACCAATTCGCGGCCGACTACCTGTGTAGGCCCG	1196
Db	1123	ACAATGCGGGAAAGATTATGAGCTGTACTAAACCGCCCGACTACCTGTGTGGCCAGC	1182
Qy	1197	CTCGCGCCGGGACATTAAAAACAACCGCATCTACCGCATCTGAGACTGAACGGCTACG	1256
Db	1183	AGCCCGTGTGAGCAATTTCAAGCAATCGAGATTTATAAAATTTGTGAACCTAAACGGGTACG	1242
Qy	1257	AACTGTGCTAAGCCGGGCTCCGTCTTTTGTGGGTGTGGGCCAGAAAAGSTTGTGGGAAGGCA	1316
Db	1243	ATCCCAATATGCGGCTCCGTCTTTTGTGGATGTGGGCCAGAAAAGSTTGTGGCAAGGGA	1302
Qy	1317	ACACCATCTGGCTTTTGGGCGGGCCACCAAGCGGCAAGCAACCATGCGCGAAAGCATCG	1376
Db	1303	ACACCATCTGGCTTTTGGGCGGTGCAACTACCGGAAGACCAACATGTGCGGAGGCCATAG	1362
Qy	1377	CCACGCGCTGTGCTCTTACGCGTGTCAACTGAGCAATGAGAACTTTCCTTCATAG	1436
Db	1363	CCCACTATGTGCTCTTCTTACGCGGTGCTAAACTGAGCAATGAGAACTTTCCTTCACG	1422
Qy	1437	ATTGTGCTGACAAAGATGTGTATCTGTGTGGAGAGGCAAGATGACCGCAAGTGTGTGG	1496
Db	1423	ACTGTGTGACAAAGATGTGTATCTGTGTGGAGAGGAGGAAAGATGACCGCAAGTGTGTGG	1482
Qy	1497	AGTCCGCGCAAGGCAATTCCTGCGCGGCGAGCAAGGTCGCTGTGACCAAAAGTGCAGTGTG	1556
Db	1483	AGTCCGCGCAAGGCAATTCCTGCGAGAGAGCAAGGTGCGCTGTGACCAAAATGCAAGTCT	1542
Qy	1557	CGGCGCAGATTCGACCCCAACCCCGTGTATGTCTCACTTCACCAACCAATGTGTGCGCGTGA	1616
Db	1543	CGGCGCAGATTCGACCCGACCTCCGTGTATGTCTCACTTCACCAACCAATGTGTGCGCGTGA	1602
Qy	1617	TTGACGGGAAACAGACCACTTGTGAGACCAAGACCGCTTGTGAGAGACCGGATGTTCAAT	1676

Db 1603 TTGACGGAACTCAACGACCTTTCGACACGACCGCTTGCAAGACCGGATGTTCAAT 1662
Qy 1677 TTGAATCAACCCCGCGTGTGGAGCATGACTTTGGCAAGTGACAAAGACAGAAAGTCAAG 1736
Db 1663 TTGAATCAACCCCGCGTGTGGAGCATGACTTTGGCAAGTGACAAAGACAGAAAGTCAAG 1722
Qy 1737 AGTTCCTTCCGCTGGGCGCAGAGATCAAGTGAACGAGTGGCGCATGAGTTTCACTCAGAA 1796
Db 1723 ACTTTTCCGGTGGCAAAAGGATCAAGTGGTGAAGTGAAGCATGATTTCTACGTCAAA 1782
Qy 1797 AGGCTGAGCCACAAAGAACCCGCCCATGACGCGGATTAAGAGCGAGCCAAAGCGGG 1856
Db 1783 AGGCTGAGCCACAAAGAACCCGCCCATGACGCGGATTAAGAGCGAGCCAAAGCGGG 1842
Qy 1857 CTTGCCCTCAGTGGCGGATTCATCGACGTCAGACGCGGAGAGCTCCGCTGAGACTTGG 1916
Db 1843 TGGCGAGTCAAGTGGCGGATTCATCGACGTCAGACGCGGAGAGCTCCGCTGAGACTTGG 1899
Qy 1917 CCGACAGGTACCAAAACAAATGTTCTGTCACGCGGCGCATGCTTCAAGATGCTTTCCCT 1976
Db 1900 CAGACAGGTACCAAAACAAATGTTCTGTCACGCGGCGCATGATGCTGATGCTTTCCCT 1959
Qy 1977 GCAAGACATGCGAGAGATGATGATGATTTCAATTTCTTCACTGACGACGGAAGAGAG 2036
Db 1960 GCAAGACATGCGAGAGATGATGATGATTTCAATTTCTTCACTGACGACGGAAGAGAG 2019
Qy 2037 ACTGTTCAAGTGTCTCCCGGCGTGTCAAGATTCGACACGGT---CGTCAAGAAAGGA 2093
Db 2020 ACTGTTCAAGTGTCTCCCGGCGTGTCAAGATTCGACACGGT---CGTGTCAAGAAAGG 2076
Qy 2094 CGTATCGGAATCTGTGCAATTCATCATCTGTGGGCGGCGCTCCGAGATTTGCTTGT 2153
Db 2077 CGTATCGGAATCTGTGCAATTCATCATCTGTGGGCGGCGCTCCGAGATTTGCTTGT 2133
Qy 2154 CGGCTGCGATCTGTGTCACGTTGACCTGATGATCTGTGTTCTGAGCAATTAATGACTT 2213
Db 2134 CTGCTGCGATCTGTGTCACGTTGATGATGATGATGATGATGATGATGATGATGATGAT 2193
Qy 2214 AACCAAGTATGCTGCGGATGATTTCTTCAAGATTTGCTGAGGACAACTCTGAG 2273
Db 2194 AATTCAGATGATGCTGCGGATGATTTCTTCAAGATTTGCTGAGGACAACTCTGAG 2253
Qy 2274 GGCATTCGAGTGTGGGACTTGAACCTGAGCCCGAAGCCCAAGCCCAACGACAA 2333
Db 2254 GGAATTAAGACAGTGTGGGACTTGAACCTGAGCCCGAAGCCCAAGCCCGGAGAGCG 2313
Qy 2334 AAGCAGACGACGCGCGGCTGTGTCTTCTGCTGCTAAGATACCTCGGACCTTTCAAC 2393
Db 2314 CATTAAGACGACGCGCGGCTGTGTGTCTTCTGCTGCTAAGATACCTCGGACCTTTCAAC 2273
Qy 2394 GGAATTAAGACGCGCGGCTGTGTGTCTTCTGCTGCTAAGATACCTCGGACCTTTCAAC 2453
Db 2374 GGAATTAAGACGCGCGGCTGTGTGTCTTCTGCTGCTAAGATACCTCGGACCTTTCAAC 2433
Qy 2454 GGCATTCGAGTGTGGGACTTGAACCTGAGCCCGAAGCCCAAGCCCAACGACAA 2513
Db 2434 GGCATTCGAGTGTGGGACTTGAACCTGAGCCCGAAGCCCAAGCCCGGAGAGCG 2493
Qy 2514 GCCGAGTTTCAAGAGCGTCTGCAAGAGATACCTTTTGGGGCAACCTCGGCGAGCA 2573
Db 2494 GCCGAGTTTCAAGAGCGTCTGCAAGAGATACCTTTTGGGGCAACCTCGGCGAGCA 2553
Qy 2574 GTCTTCAGGCGCAAGAGCGGCTTCTGCAAGCTCTGGCTGTGTTGAAGAGCGCTTAAG 2633
Db 2554 GTCTTCAGGCGCAAGAGCGGCTTCTGCAAGCTCTGGCTGTGTTGAAGAGCGCTTAAG 2613
Qy 2634 ACGGCTCTTGAAGAAACGTCGATGAGACGTCGACGACGACGACGACGACGACGAC 2693
Db 2614 ACGGCTCTTGAAGAAACGTCGATGAGACGTCGACGACGACGACGACGACGACGAC 2673
Qy 2694 GGCATGCGCAAGACGCGCGACGCGCTGTAAGAAAGACTCAATTTTGTGACGCTGCG 2753
Db 2674 GGCATGCGCAAGACGCGCGACGCGCTGTAAGAAAGACTCAATTTTGTGACGCTGGA 2733

Qy 2754 GACTCAGAGTCAAGTCCCGATTCACAACTCTTGGAGAACTTCAGACACCCCGCTGCT 2813
Db 2734 GAGCAGACTCAAGTCAAGTCCCGATTCACAACTCTTGGAGAACTTCAGACACCCCGCTGCT 2793
Qy 2814 GTGGGACCTTCAATGAGCTTCAAGCGGCTGGCGACCAATGAGCAGACAAATACGAGGCG 2873
Db 2794 CTGGGACCTTCAATGAGCTTCAAGCGGCTGGCGACCAATGAGCAGACAAATACGAGGCG 2853
Qy 2874 GCCGACGAGTGGGATATGCTCAGGAAATTTGGCAATTTGCAATGCTGAGGCGAGC 2933
Db 2854 GCCGACGAGTGGGATATGCTCAGGAAATTTGGCAATTTGCAATGCTGAGGCGAGC 2913
Qy 2934 AAGATCATACCAACGACGACCCGACACTGGGCTGGCCCACTAATTAACCACTCTAC 2993
Db 2914 AAGATCATACCAACGACGACCCGACACTGGGCTGGCCCACTAATTAACCACTCTAC 2973
Qy 2994 AAGCAATCTCCAGTGTCTTCAACGCGGCGCAGACCAACCACTACTTGGCTACAGC 3053
Db 2974 AAGCAATCTCCAGTGTCTTCAACGCGGCGCAGACCAACCACTACTTGGCTACAGC 3030
Qy 3054 ACCCTTGGGCGTATTTGATTTCAACAGATTCACCTGCACTTTTACACGCTGAC 3113
Db 3031 ACCCTTGGGCGTATTTGATTTCAACAGATTCACCTGCACTTTTACACGCTGAC 3090
Qy 3114 CAGGACCTCATCAACAAATTTGGGATTCGCGCCCAAGACTCACTTCAACTCTTC 3173
Db 3091 CAGGACCTCATCAACAAATTTGGGATTCGCGCCCAAGACTCACTTCAACTCTTC 3150
Qy 3174 AACATCAAGTCAAGAGGCTACAGACGAATGATGCGTCAACCACTCTGATTAACCTT 3233
Db 3151 AACATCAAGTCAAGAGGCTACAGACGAATGATGCGTCAACCACTCTGATTAACCTT 3210
Qy 3234 ACCAGACGCTTCAAGTCTTCTGGAATGAGTACAGCTTCCGATCTGCTGCTGCT 3293
Db 3211 ACCAGACGCTTCAAGTCTTCTGGAATGAGTACAGCTTCCGATCTGCTGCTGCTGCT 3270
Qy 3294 GCGCACGAGGCGCTCCCTCGCTTCCGCGGAGAGCTGATTAATTCGGAATACGCG 3353
Db 3271 GCGCACGAGGCGCTCCCTCGCTTCCGCGGAGAGCTGATTAATTCGGAATACGCG 3330
Qy 3354 TACCTGACGCTCAACAAATGAGCAGCAAGCGCTGGGACGTTCACTTCACTGCTGGA 3413
Db 3331 TACCTGACGCTCAACAAATGAGCAGCAAGCGCTGGGACGTTCACTTCACTGCTGGA 3390
Qy 3414 TATTTCTCTTCAAGTGTGAGAACGCGCAACCTTTACCTTCACTGCTGGA 3473
Db 3391 TATTTCTCTTCAAGTGTGAGAACGCGCAACCTTTACCTTCACTGCTGGA 3450
Qy 3474 GAAAGTCTTTCACAGACGCTACGCGACACGCGACGCTGAGACCGCTGATGATCT 3533
Db 3451 GAAAGTCTTTCACAGACGCTACGCGACACGCGACGCTGAGACCGCTGATGATCT 3510
Qy 3534 CTCATGACCAATTAACCTGATTAACCTGAAACAGAACTCAAAATCAGTCGGAAGTGGCC 3593
Db 3511 CTCATGACCAATTAACCTGATTAACCTGAAACAGAACTCAAAATCAGTCGGAAGTGGCC 3570
Qy 3594 AACAGGACTGCTGTTTACCGTGGCTTCAAGCTGACGCTGTTCAAGCCCAAAAAC 3653
Db 3571 AACAGGACTGCTGTTTACCGTGGCTTCAAGCTGACGCTGTTCAAGCCCAAAAAC 3630
Qy 3654 TGGCTACCTGAGACCTGTTTATGCGAGACGCGCTTCTTAAACAAAACAGACAAAC 3713
Db 3631 TGGCTACCTGAGACCTGTTTATGCGAGACGCGCTTCTTAAACAAAACAGACAAAC 3690
Qy 3714 AACAGCAATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGAGCGTGAATTCATC 3773
Db 3691 AACAGCAATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGAGCGTGAATTCATC 3750
Qy 3774 ATCAACCTTGGGACCTGATGAGCTTCAACAAAGACGAAAGCAAGTTCTTCCATG 3833
Db 3751 ATCAACCTTGGGACCTGATGAGCTTCAACAAAGACGAAAGCAAGTTCTTCCATG 3810

```

QY 3834 AGCGGTGATGATTTTGGAAAAGAGCGCGAGCTTCAACACTGATGGACAAT 3893
DB 3811 AGCGGGTTCTACTTTTGGAGACGAGCTCAAGAAAACAAATGCGACATTGAAA 3870
QY 3894 GTCATGATTAAGACGAGAGAAATTAAAGCACTAACCTGTGGCCACCGAAATTT 3953
DB 3871 GTCATGATTAAGACGAGAGAAATCAAGCAACCAATCCCGTACGAGCAATAT 3930
QY 3954 GGGACCGTGGACATTTTCCAGAGCAGCAGACACCTCGGACCGGAGATGTCAT 4013
DB 3931 GGTTCCTATCTTACCACTTCCAGAGGCAACAGACAGGATACCGCAGATGTCAC 3990
QY 4014 GCTATGAGACATTAACCTGAGTGTGGAGATAGAGAGTATCTGACGGTCCC 4073
DB 3991 ACACAGGCGTTCTTCCAGGATGTGTGGAGAGAGAGATGTATCTTCAAGGGCCC 4050
QY 4074 ATTTGGGCAAAATTCCTACACAGATGACATTTTCAACCCGTCTCTTATGGCGGC 4133
DB 4051 ATCTGGGCAAAATTTCCACACAGAGGACATTTTCAACCCCTCTCCCTCATGGTGA 4110
QY 4134 TTTGAGACTCAAGAACCGCGCTCTCCATCTCATCAAAAACGGCTGTTCCTGGCAAT 4193
DB 4111 TTGCGACTTAAACCCCTCTCTCCACAGATTCATCAAGAACACCCGCTATCTGCAAT 4170
QY 4194 CCTCGGCGAGATTTTCAAGTATGCTTCAATTCATCAACCAATCTCCACAGAG 4253
DB 4171 CCTTCAGACACCTTCAAGTGGCGCAAAAGTTGCTTCTTCAATCAACAGTACTCCAGGGA 4230
QY 4254 CAAGTATGATGAAATTTGAATGAGAGTGCAGAAAGAAACAGCAAGCGTGGATCCC 4313
DB 4231 CAGTATGAGCGGAGATGAGAGTGCAGAGAGGAAACAGCAACCGTGGATCCC 4290
QY 4314 GAAATGAGTACATCAATTAATGCAAAATCTGCAACGTTGATTTTATCTGAGCAAC 4373
DB 4291 GAAATGAGTACATCTTCACTCAACAGTCTGTATGAGCTTTACTGTGAGCACT 4350
QY 4374 AATGAGCTTATATCTGAGCTCGGCCCAATGAGCAACCGCTTACTTACCCGCTGAA 4433
DB 4351 AATGCGGTATTAAGAGCGCTCGGCCCAATGAGCAACGATACCTGATCTGATTA 4410
QY 4434 TTACGTGATTAATCAATTAACCGGTTGATGCTTCACTGAGTAACTTTGCTCTGCTT 4493
DB 4411 TTGCTGTATTAATCAATTAACCGGTTGATGCTTCACTGAGTAACTTTGCTCTGCTT 4470
QY 4494 CTATCTTATGATTAACGATGATTAAGCTTACATTAATCTGCTGAGTTCGCTGCG 4553
DB 4471 TCTTCTTATTAATGATTCATGCTGCTAGTATTAAGTATGAGCGGCTTATATTAAC 4530
QY 4554 ATAAAGACTTAACGATCGGCTTACCCCTAGTATGATGATGCGCACTCCCTCTGCG 4613
DB 4531 TACAAGG-----ACCCCTAGTATGATGATGCGCACTCCCTCTGCG 4574
QY 4614 CGCTGCTGCTCGGTGGGCTTGGGACCAAGGTTCGCGACAGGAGAGTCTGCTCT 4673
DB 4575 CGCTGCTGCTGCTGCTGAGGCGGCGACCAAGGTTCGCGCGCGGCTGCTGCGG 4634
QY 4674 GCGGCGCGCGACCGAGCGAGCGCGGCGAGAGAGGAGTGGGCA 4718
DB 4635 GCGGCGCTGAGTGAAGCGAGCGCGGCGAGAGAGGAGTGGGCA 4679

```

RESULT 3

```

US-10-038-972A-12
; Sequence 12, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05

```

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-038-972A-12

Query Match      65.9%; Score 3107.8; DB 15; Length 4679;
Best Local Similarity 80.2%; Pred. No. 0;
Matches 3790; Conservative 0; Mismatches 882; Indels 53; Gaps 10;

1 TTGCGCACTCCCTCTCTGCGGCTCGCTGCTGAGTGGGCTTGCAGCAAGTCCG 60
1 TTGGGCACTCCCTCTCTGCGGCTCGCTGCTGAGTGGGCTTGCAGCAAGTCCG 60
QY 61 AGAGCGAGAGCTCTGCTCTGCGGCTTGCAGGAGGAGGAGGAGGAGGAGG 120
DB 61 CGAGCGCGGCGCTTTCGCGGCGGCTTGCAGGAGGAGGAGGAGGAGGAGG 120
QY 121 GCGAATCCATCACTAGGAGGATTCGGAAGCGCTTCCACGCTGCGCGCTGA 180
DB 121 GCGAATCCATCACTAGGAGGATTCGGAAGCGCTTCCACGCTGCGCGCTGA 180
QY 181 CGTAAATTAAGTATAGG---GAGTGTCTGTATTAAGTGTACGTGAGTCTTTTG 237
DB 164 CGTAAATTAAGTATAGGAGGATTCGGAAGGAGGAGGAGGAGGAGGAGGAG 222
QY 238 GATATTTTGGAGACCAAGTGGCTTATTTAGGCTTATTTAGGCTTATTTAG 297
DB 223 GATATTTTGGAGACCAAGTGGCTTATTTAGGCTTATTTAGGCTTATTTAG 282
QY 298 CTCATTTTGG-ACCGGAAATTTGAACGAGCAGCAGCAGCAGCAGCAGCAG 356
DB 283 CTCATTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 342
QY 357 TGATCAAGGTGCGAGCGACCTGAGCAGCAGCAGCAGCAGCAGCAGCAGC 416
DB 343 TGATCAAGGTGCGAGCGACCTTGAAGCAGCAGCAGCAGCAGCAGCAGCAG 402
QY 417 GCTGGGTGGCGGAGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 476
DB 403 ACTGGGTGGCGGAGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 462
QY 477 TTGAGCAGGACCGCTGAGCGGTGCGGAGAGTGGAGTGGAGTGGAGTGGAG 536
DB 463 TTGAGCAGGACCGCTGAGCGGTGCGGAGAGTGGAGTGGAGTGGAGTGGAG 522
QY 537 GCGCGGTGAGTAAGCGCGGAGGCGCTTCTTTGTTCAAGTTCGAGAGGCGAG 596
DB 523 GCGGTGAGTAAGCGCGGAGGCGCTTCTTTGTTGCAATTTGAGAGGAGAGGT 582
QY 597 ACTTCCACCTCCATATTTGTTGAGAGCAGCGGAGGAGTGGAGTGGAGTGG 656
DB 583 ACTTCCACAGCAGCGGTGTTGAGAGCAGCGGAGGAGTGGAGTGGAGTGG 642
QY 657 TCTGAGTCAAGTTAGGAGCAAGCTGTGAGCAGCAGTTCACGCGGAGTTCAG 716
DB 643 TCTGAGTCAAGTTAGGAGCAAGCTGTGAGCAGCAGTTCACGCGGAGTTCAG 702
QY 717 TGCCCAACTGGTTGCGGTGAGCAAGACCGGTATGCGCGGAGGAGGAGGAG 776
DB 703 TGCCCAACTGGTTGCGGTGAGCAAGACCGGTATGCGCGGAGGAGGAGGAG 762
QY 777 TGAGCAGAGTCAATTCCTCACTCTGCGGAGAGTTCAGCGCGGAGTTCAG 836
DB 763 TGAGCAGAGTCAATTCCTCACTCTGCGGAGAGTTCAGCGCGGAGTTCAG 822
QY 837 CGTGAATTAATGAGAGGAGTATATAGGCGCTGTTTGAACCTGCGCGAGCGCA 896
DB 823 CGTGAATTAATGAGAGGAGTATATAGGCGCTGTTTGAACCTGCGCGAGCG 882
QY 897 TCGTGGCGAGCACTGAGCCAGCTGAGCCAGGAGCAGGAGGAGGAGGAGTCTGA 956

```

```
Db      883 TGTGGCGCAGACTCTGACGCACTGTGCGCAGACGACGAGCAGAAACAAAGAAATCAGA 942
Qy      957 ACCCCATTTCTGAGCGCGCTGTACTCGGTGTAACAACTCCGCGCGCTCACTGAGACTGG 1016
Db      943 ATCCCAATTTCTGAGCGCGGTGTATCAGATCAAAAATCTTCAGCCAGGTATACATGAGCTGG 1002
Qy      1017 TCGGGTGGCTGTGAGACCGGGGGCTCACTCCGAGAAAGAGTGGATCCAGAGGACGAG 1076
Db      1003 TCGGGTGGCTGTGAGACAGGGGATTTACTGTGGAAAGCAGTGGATCCAGAGGACGAG 1062
Qy      1077 CCTGTGATCTCTCTTCAATGCGCGCTCAATCTGCGGCTCCCAATCAAGGCTGCTTGG 1122
Db      1063 CCTCATCATCTCTCTTCAATGCGCGCTCAATCTGCGGCTCCCAATCAAGGCTGCTTGG 1122
Qy      1137 ACAATGCCGCGAGATCATGCGCTGACCAATCCGCGCGCTCACTCTGTAGGCCCCG 1196
Db      1123 ACAATGGGGGAAATATATAGCTGACTTAAACCGCGCGCTCACTCTGTAGGCCCCG 1182
Qy      1197 CTCCGCGCGCGGACATTAAACCAACCGCATCTACCGCATCTGTAGGCTGAGCGCTAGC 1256
Db      1183 AGCCCGTGGAGGACATTTCCAGCAATCGAATTTTAAATTTGAACTAAACGAGTACG 1242
Qy      1257 AACCTGCTTACGCGCGCTCCGTCTTTCTCGGCTGGGCGCAGAAAAGTTGCGGAGCGCA 1316
Db      1243 ATCCCAATATAGCGCTTCCGTCTTTCTGGGATGGGCGAGAAAAGTTGCGGAGAGGA 1302
Qy      1317 ACACCATCTGCTTTTGGGCGCGGCGCACACGCGGACAGACACATGCGGAGACCATG 1376
Db      1303 ACACCATCTGCTTTTGGGCGCTGCAACTACCGGGAAGACCAACATGCGGAGCGCATAG 1362
Qy      1377 CCCAGCGCGTGCCTTCAAGCGGCTGGGTCACATGAGCAATGAGAACTTTCCCTTCAATG 1436
Db      1363 CCCAGCTGTGCTCTTCAAGGCTGGGTCAGTAATGAGCAATGAGAACTTTCCCTTCAAG 1422
Qy      1437 ATTCGCTGCAAGATGATGATCTGTGAGAGAGGCAAGATGACCGCCAAAGGTCGTG 1496
Db      1423 ACTGTGTCGACAAAGATGATGATCTGTGAGAGAGGGAAGATGACCGCCAAAGGTCGTG 1482
Qy      1497 AGTCCGCGAAGGCGCTTCTGCGCGGCGACAGAAAGTGGCGGTGAGCAAAAGTCAAGTGT 1556
Db      1483 AGTCCGCGAAGGCGCTTCTGCGCGGCGACAGAAAGTGGCGGTGAGCAAAAGTCAAGTGT 1542
Qy      1557 CCGCGCAGATGACCCGACCGCGGTGATGTCATCTTCAACCAACCAATGCGCGCGTGA 1616
Db      1543 CCGCGCAGATGACCCGACCTCGGTGATGTCATCTTCAACCAACCAATGCGCGCGTGA 1602
Qy      1617 TTGACGGGAACAGACACCACTTTCAGAGACAGACGCGCTTTCAGAGACCGGATGTTCAAT 1676
Db      1603 TTGACGGGAACAGACACCTTTCAGAGACAGACGCGCTTTCAGAGACCGGATGTTCAAT 1662
Qy      1677 TTGACCTCACCCCGCTGTGAGAGATGACTTTGGCAAGGTGACAAAGAGAGAGTCAAG 1736
Db      1663 TTGAACTCACCCCGCTGTGATGATGACTTTGGGAAGGTGACCAAGAGAGAGTCAAG 1722
Qy      1737 AGTTCTTCCGCTGGGCGCAGATGACGTGACCGAGGTGGGAGATTTTAAAGTCAAG 1796
Db      1723 ACTTTTCCGCTGGGCGCAGAGATGACGTGTTGAGGTGAGACATGAAATTTCTACGTCAAA 1782
Qy      1797 AGGCTGAGGCAACAAAGACCCGCGCGATGACGCGGATTTAAAGCGAGCCCAAGCGGG 1856
Db      1783 AGGCTGAGGCAACAAAGACCCGCGCGATGACGCGGATTTAAAGCGAGCCCAAGCGGG 1842
Qy      1857 CTGCGCTTCACTGCGGATTCATGACGTGACGCGGAGAGAGTCCGCTGAGACTTTG 1916
Db      1843 TGGCGAGTCACTGCGGATTCATGACGTGACGCGGAGAGAGTCCGCTGAGACTTTG 1899
Qy      1917 CCGCAGAGTCAACAAAGATGTTCTGTGACGCGGCGGATGCTTCAAGTCTGTTTCCCT 1976
Db      1900 CAGCAGAGTCAACAAAGATGTTCTGTGACGCGGCGGATGAACTGATCTGTTTCCCT 1959
Qy      1977 GCAAGATCGAGAGATGATCAGAAATTTCAACTTTGCTTCAACGACGCGGAG 2036
Db      1960 GCAGACAATCGAGAGATGATCAGAAATTTCAATATCTGCTTCACTCAACGACAGAAAG 2019
Qy      2037 ACTGTTCAAGATGCTTCCCGGCTGTGAGAAATCTCAACCGGT---GTCAGAAAGAGA 2093
Db      2020 ACTGTTTAAAGTCTTTC---GTCAGAAATCTCAACCGGTTCCTGTCGCAAAAAG 2076
Qy      2094 CGATCGGAAACTCTGTGCAATTCATCTGTGCGGGCGGGCTCCCGAGATGCTTGGCT 2153
Db      2077 CGATCGGAAAATCTGTGCTATCTATCTATCTATGAGAAAAGGTGCCAG---CGCTTGA 2133
Qy      2154 CCGCTCGCATCTGTGCAAGTGAACCTGATGATCTGTGTTCTGACCAATTAATGACTT 2213
Db      2134 CTGCTGCGCATCTGTGCAATGATGATTTGGATGATGATCTTGTGAACATTAATGATTT 2193
Qy      2214 AATCAGATATGCTGCTCCGATGATTTCTTCCAGATTTGGCTGAGAGACAACTCTGAG 2273
Db      2194 AATCAGATATGCTGCTCCGATGATTTCTTCCAGATTTGGCTGAGAGACAACTCTGGA 2253
Qy      2274 GGCATTCGCGAGTGTGGGACTTGAACCTGAGAGCCCGAAGCGCAACGACGAA 2333
Db      2254 GGAATTAAGACAGTGTGAGAGCTCAACCTGCGCCACACACCAAGCCCGAGAGCGG 2313
Qy      2334 AAGCAGAGACGCGCGGCTGTGATGCTTCTGTGCTAAGATACCTGAGACCTTTCAAC 2393
Db      2314 CATAGAGACGACAGAGGGTCTTGTGCTTCTGGGTACAAAGTACCTCGGACCTTTCAAC 2373
Qy      2394 GAACTGACAGAGGGGAGCCCTGCAACGCGCGGACGACAGCGGCTCTGAGACGACAA 2453
Db      2374 GAACTGACAGAGGAGCGCGTCAACGAGGACAGCGCGGCTCTGAGACGACAA 2433
Qy      2454 GCTTACGACAGAGGCTCAAGCGGGTGAACAATCGGTAACCTGCGATTAACGCGCGAC 2513
Db      2434 GCTTACGACCGGAGCTTCAAGCGGAGACACCGGTAACCTGCGATTAACGCGCGAC 2493
Qy      2514 GCCAGTTTCAGAGCGCTGTGCAAGAAATACGCTTTTGGGGGCAACCTCGGGCGAGCA 2573
Db      2494 GCGAGTTTCAGAGAGGCTTAAAGAAATACGCTTTTGGGGGCAACCTCGGGCGAGCA 2553
Qy      2574 GTTTTCAGGCGCAAGAGGGGTTCTGAACTCTCGGTGCTGAGAGAGGCGCTAAG 2633
Db      2554 GTTTTCAGGCGCAAGAGGGGTTCTGAACTCTCGGTGCTGAGAGAGGCGCTAAG 2613
Qy      2634 AGGCTCTGGAAGAAAGTCCGCTGAGAGAGTGGCGCAAGAGGCAAGTCTCTG 2693
Db      2614 AGGCTCTGGAAGAAAGTCCGCTGAGAGAGTGGCGCAAGAGGCAAGTCTCTG 2673
Qy      2694 GGCATCGGAGAGCAGGCGCAGAGCGCGCTAAAGAGACTCAATTTGCTGAGCTGAG 2753
Db      2674 GGAACCGGAAAGCGCGGCGCAGAGCGCTGCAAGAAAGATTTGCTGAGCTGAG 2733
Qy      2754 GACTCAAGTCACTGCTCCGATTCACAACTCTGCGAAGAACTTCAGCAACCCCGCTGCT 2813
Db      2734 GAGCGAGCTCACTGCTGCTGAGAGAGTGGCGCAAGAGGCAAGTCTCTG 2793
Qy      2814 GTGGGACTTACTCAATATGCTTCAAGCGGCTGCGGCGCAATGAGGCAAGTAAAGAGG 2873
Db      2794 CTGGGAACTTAATGAGTCTCAAGCAAGTGGGCGCAATGAGGCAAGTAAAGAGG 2853
Qy      2874 GCCGAGAGTGGGTATGCTCAGAGAAATTTGGCATTTGAGATTCACATGCTGAGCGAG 2933
Db      2854 GCCGAGAGTGGGTATGCTCAGAGAAATTTGGCATTTGAGATTCACATGAGTGGCGAG 2913
Qy      2934 AGAGTATATCAACACGACCGGACCTGCGGCTTGGCCACTTACATTAACCACTTAC 2993
Db      2914 AGAGTATATCAACACGACCGGACCTGCGGCTTGGCCACTTACATTAACCACTTAC 2973
Qy      2994 AAGCAATCTCCAGTCTTCAACGCGGCGCAGAGAGCAACCACTTCTGCTAAG 3053
Db      2974 AAGCAATCTCCAGCAATCA---GAGGCTTCAAGCAAGTCACTTCTGCTAAG 3030
Qy      3054 ACCCGCTGGGGATTTGATTTCAACAGATTTCACTGCACTTTTCAACAGTGAAG 3113
Db      3031 ACCCGTTGGGGATTTTGACTTCAACAGATTTCACTGCACTTTTCAACAGTGAAG 3090
```

QY 3114 CAGGACTCATCAACAATTTGGGGATTCCGGCCCAAGAGACTCAACTTCAACTCTTC 3173
 Db 3091 CAAGACTCATCAACAACCTGGGGATTCCGAGCCAGAGACTCAACTTCAACTCTTC 3150
 QY 3174 AACATCAAGTCAAGAGTCAAGAGTCAAGATGAGCTTCAACCACTCGCTAATACCTT 3233
 Db 3151 AACATCAAGTCAAGAGTCAAGAGTCAAGATGAGCTTCAACCACTCGCTAATACCTT 3210
 QY 3234 ACCGACGCTTCAAGTCTTCCGAGCTCGAGTACCAAGCTTCCGATCTCTGGCTCT 3293
 Db 3211 ACCGACGCTTCAAGTCTTCCGAGCTCGAGTACCAAGCTTCCGATCTCTGGCTCT 3270
 QY 3294 GCGACACAGGCTGCTCCCTCCGCTCCGAGAGCTGCTAATGATTCGCAATACGAC 3353
 Db 3271 GCGACATCAAGAGTCTTCCGCTCCGAGAGAGCTTCAAGTCTTCAAGTCTTCAAG 3330
 QY 3354 TACCTGACGCTCAACATGAGAGCCAGCCGCTGAGAGCTTCACTTCTGCTGAG 3413
 Db 3331 TACCTGACGCTCAACATGAGAGCCAGCCGCTGAGAGCTTCACTTCTGCTGAG 3390
 QY 3414 TATTTCCCTTCTGAGTCTGAGAAACGGGCAACATTTACTTCACTTCACTTCTGAG 3473
 Db 3391 TACTTCTCTCTGAGTCTGAGAAACGGGCAACATTTACTTCACTTCACTTCTGAG 3450
 QY 3474 GAGGCTCTTCTCAAGAGCTGAGAGCCAGCCGCTGAGAGCTTCACTTCTGAGTCT 3533
 Db 3451 GAGCTTCTTCTCAAGAGCTGAGAGCCAGCCGCTGAGAGCTTCACTTCTGAGTCT 3510
 QY 3534 CTCACTGACCAATCTGATTAATCTGAAAGAACTCAATCAATCAATCAATCAATCAAT 3593
 Db 3511 CTCACTGACCAATCTGATTAATCTGAAAGAACTCAATCAATCAATCAATCAATCAAT 3570
 QY 3594 AACAGAGCTTCTGAGTCTGAGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 3653
 Db 3571 CAGTCAAGCTTCAAGTCTGAGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 3630
 QY 3654 TGGCTACCTGAGACCTGTTATCGGAGAGAGCCGCTTCAACAAACAAACAGACACAC 3713
 Db 3631 TGGCTTCTGAGACCTGTTATCGGAGAGAGAGTCAATCAATCAATCAATCAATCAAT 3690
 QY 3714 AACAGCAATTTTACCTGAGTCTGAGTCTTCAATCAATCAATCAATCAATCAATCAAT 3773
 Db 3691 AACAGTGAATCTGAGTCTGAGTCTGAGTCTTCAATCAATCAATCAATCAATCAATCAAT 3750
 QY 3774 ATCAACCTGAGACCTGTTATCGGAGAGAGCCGCTTCAACAAACAAACAGACACAC 3833
 Db 3751 GTGATTCGGGCTCGGCTGAGAGAGCCAGAGAGAGTCAATCAATCAATCAATCAATCAAT 3810
 QY 3834 AGCGGTCTCATGATTTTGGAAAAGAGAGCCGCTGAGTCAATCAATCAATCAATCAATCAAT 3893
 Db 3811 AGCGGGTCTCATGTTTGGAAAAGAGAGCTTCAAGAAAACAAATGTCGATGAAAG 3870
 QY 3894 GTCATGATTAACAGAGAGAGAAATTAAGCCACTAATCCGTGAGCAACGAGAAATTT 3953
 Db 3871 GTCATGATTAACAGAGAGAGAAATTAAGCCACTAATCCGTGAGCAACGAGAGAT 3930
 QY 3954 GGGACCGGAGAGTCAATTTCCAGAGAGAGAGCCGCTGAGAGAGAGTGTGAT 4013
 Db 3931 GGTCTGATTAACAG 3990
 QY 4014 GCTATGAGAGATTAACCTGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4073
 Db 3991 ACAGAGAGAGTCTTCTTCCAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 4050
 QY 4074 ATTTGGGCAAAATTTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4133
 Db 4051 ATTTGGGCAAAATTTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4110
 QY 4134 TTTGAG 4193
 Db 4111 TTTGAG 4170

QY 4194 CTTCCGCGAGGTTTTCAGTCAACAAGTTTGTCTTCAATTCATCCCAATATCTCAGAGA 4253
 Db 4171 CTTTCAGACACCTTCAAGTGTGGGAGAGAGTTGCTTCTTATATCAAGATATCTCAGAGA 4230
 QY 4254 CAGTGAAGTGTGAATTTGATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4313
 Db 4231 CAGTGAAGTGTGAGAGTGTGAG 4290
 QY 4314 GAGTGAAGTGTGAGAGTGTGAG 4373
 Db 4291 GAGTGAAGTGTGAGAGTGTGAG 4350
 QY 4374 AATGACTTATATGAGAGCTTCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4433
 Db 4351 AATGAGGAGTATTCAG 4410
 QY 4434 TTAGGTTTATCAATTAACAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4493
 Db 4411 TTAGGTTTATCAATTAACAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4470
 QY 4494 CTTATCTTATCGGTTTCAATGATTTATAGCTTACATTAATCACTTGTGAGAGAGAGAGAG 4553
 Db 4471 TCTTTCTTATCTAGTTTCAATGAGCTTACATGATTAATGATGAGAGAGAGAGAGAGAGAG 4530
 QY 4554 ATAAAGACTTACGATCGAGTTTACCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4613
 Db 4531 TACAGAGA-----ACCCCTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4574
 QY 4614 CGCTCGCTCGCTGAG 4673
 Db 4575 CGCTCGCTCGCTGAG 4634
 QY 4674 GCGGCGCCACGAG 4718
 Db 4635 GCGGCGCTCAGTGAAG 4679

RESULT 4

US-10-077-294-1

: Sequence 1, Application US/10077294

: Publication No. US2002015979A1

: GENERAL INFORMATION:

: APPLICANT: Johnson, Phillip R.

: TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods

: NUMBER OF SEQUENCES: 3

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

: STREET: 6300 Sears Tower, 233 S. Wacker Drive

: CITY: Chicago

: STATE: Illinois

: COUNTRY: USA

: ZIP: 60606

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: IBM PC compatible

: SOFTWARE: Patent in Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/077,294

: FILING DATE: 15-Feb-2002

: CLASSIFICATION: <Unknown>

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: 09/691,604

: FILING DATE: <Unknown>

: ATTORNEY/AGENT INFORMATION:

: NAME: No. US2002015979A1and, Greta E.

: REGISTRATION NUMBER: 35,302

: REFERENCE/DOCKET NUMBER: 31975

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (312) 474-6300

: TELEFAX: (312) 474-0448

: TELETYPE: 25-3856

1977 GCACAGATGCGAGAGATGAATCAGATTTCACATTTGGCTTCAGCGACGGACGAG 2036
1960 GCACAGATGCGAGAGATGAATCAGATTTCACATTTGGCTTCAGCGACGGACGAG 2019
2037 ACTGTTCAAGAGTGTCTCCCGCGCGTGTCTAGATCTCAACCGGT---CGTCAGAAAGAGA 2093
2020 ACTGTTAGAGTGTCTCC---CGTGTCAAGATCTCAACCGGTTCGTGTCTCAAAAGG 2076
2094 CGTATCGAAACTCTGTGCTTTCATCATCTGCTGGGGGGGCTCCCGAGATGCTTGTCT 2153
2077 CGTATCGAAACTCTGTGCTTTCATCATCTGCTGGGGGGGCTCCCGAGATGCTTGTCT 2133
2154 CGGCTCGAGATCTGTCAACGTGACCTGATGATGATGATGATGATGATGATGATGATGAT 2213
2134 CTGCTCGAGATCTGTCAACGTGATGATGATGATGATGATGATGATGATGATGATGAT 2193
2114 AAACAGAGTATGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2273
2194 AAATCAGATATGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2253
2274 GGCATTTGCGAGTGTGAGATTTGAAACTGTGAGCCCGGAAAGCCCAAGCCACAGCA 2333
2254 GGAATTAAGACAGTGTGAGATCTCAACCTGCGCCACACCAACCAAGCCCGAGACGG 2313
2234 AAGCAGAGCAGCGCGCGGATCTGTGCTCTGTGCTCAAGTACCTGAGACCTTCAAC 2393
2214 CATAGAGACAGCAGAGGGGCTTGTGTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2373
2294 GGCATTCAGCAAGGGGGGCGCGTCAACCGCGGAGACGCGCGCGCTCGAGCAGACAA 2453
2274 GGCATTCAGCAAGGGGGGCGCGTCAACCGCGGAGACGCGCGCGCTCGAGCAGACAA 2433
2254 GGCATTCAGCAAGGGGGGCGCGTCAACCGCGGAGACGCGCGCGCTCGAGCAGACAA 2513
2234 GGCATTCAGCAAGGGGGGCGCGTCAACCGCGGAGACGCGCGCGCTCGAGCAGACAA 2493
2214 GGCAGATTCAGAGCGCTCTGACAGCGAGACCAACCGTCAAGTACCAACCAAGCCGAC 2573
2194 GCGAGATTCAGAGCGCTCTGACAGCGAGACCAACCGTCAAGTACCAACCAAGCCGAC 2553
2174 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2633
2154 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2613
2134 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2693
2114 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2673
2094 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2753
2074 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2733
2054 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2813
2034 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2833
2014 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2873
1994 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2853
1974 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2933
1954 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2913
1934 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2993
1914 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 2973
1894 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 3053
1874 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 3030
1854 GTCCTCAGGCGCAAGAGCGGGTCTCTGACCTCTCGGTGTGTGTGTGTGTGTGTGTGTGT 3113

3031 ACCCTTTGGGGGATTTTGACTTTCACAGATTCACATGCACTTTTACACAGTGAAG 3090
3114 CAGGATCTATCAACCAAACTTTGGGATTTCCGGCCCAAGAGATCTCAACTTTTCACTTTC 3173
3091 CAAAGATCTATCAACCAAACTTTGGGATTTCCGGCCCAAGAGATCTCAACTTTTCACTTTC 3150
3174 AACATCAAGTCAAGAGGATCAGACAGATGATGAGGCTCAACCACTTGTGATTAACCTT 3233
3151 AACATCAAGTCAAGAGGATCAGACAGATGATGAGGCTCAACCACTTGTGATTAACCTT 3210
3234 ACCAGACGGTTCAAGTCTTCTGGAATGAGATGAGATGAGATGAGATGAGATGAGATGAG 3293
3211 ACCAGACGGTTCAAGTCTTCTGGAATGAGATGAGATGAGATGAGATGAGATGAGATGAG 3270
3294 GCGCACAAGGATGCT 3353
3271 GCGCACAAGGATGCT 3330
3354 TACCTGAGCTCAACATGAGCAGACCGAGCGTGGAGCTTCATCTTTTACTGCTGAA 3413
3331 TACCTGAGCTCAACATGAGCAGACCGAGCGTGGAGCTTCATCTTTTACTGCTGAA 3390
3414 TATTTCTCTCTCTCAAGTCTGAGAACGGGCAACACTTTTACTGCTGAACTTTTGA 3473
3391 TATTTCTCTCTCTCAAGTCTGAGAACGGGCAACACTTTTACTGCTGAACTTTTGA 3450
3474 GAAGTCTCTTCAACAGAGCTTACGGGCAACGGGCAACGGGCAACGGGCAACGGGCAACGG 3533
3451 GAAGTCTCTTCAACAGAGCTTACGGGCAACGGGCAACGGGCAACGGGCAACGGGCAACGG 3510
3534 CTCATGACCAATATCTGATTTACTGTAACAGAACTCAAAATGATGCTCGGAAGTGGCCAA 3593
3511 CTCATGACCAATATCTGATTTACTGTAACAGAACTCAAAATGATGCTCGGAAGTGGCCAA 3570
3594 AACAGAGCTTGTGTTTAAAGCGGTGAGCTTCAAGTGGAGTGTCTGTTCAGCCCAAAAC 3653
3571 CAGTCAAGGCTTCAAGTGTCTTCAAGCGGAGGAGTGAATTCGGGACCACTGTAGGAAAC 3630
3654 TGGCTACTGAGACCTCTTATTCGGGAGGAGGCTTCTTAAACCAAAACCAAGCAAC 3713
3631 TGGCTACTGAGACCTCTTATTCGGGAGGAGGCTTCTTAAACCAAAACCAAGCAAC 3690
3714 AACGCAATTTTACCTGAGTGTGCTTCAAAATATTAACCTCAATGAGGCTGATCATC 3773
3691 AACGCAATTTTACCTGAGTGTGCTTCAAAATATTAACCTCAATGAGGCTGATCATC 3750
3774 ATCAACCTTGGACCTGATGCTTCAACCAAGAGAGAGCAAGATTTCTTCCATG 3833
3751 ATCAACCTTGGACCTGATGCTTCAACCAAGAGAGAGCAAGATTTCTTCCATG 3810
3834 AGCGGTGTATGATTTTGTGAAAGAGAGCGCGGAGCTTCAACCACTGATTTGACAT 3893
3811 AGCGGTGTGTATTTTGTGAAAGAGAGCGCGGAGCTTCAACCACTGATTTGACAT 3870
3894 GTCATGATTCACAG 3953
3871 GTCATGATTCACAG 3930
3954 GGGACGCTGAGCTCAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4013
3931 GGGACGCTGAGCTCAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3990
4014 GTCATGAGATTCATCTGAG 4073
3991 ACACAAAGCTTCTTCAAG 4050
4074 ATTGGGCAAAATTTCTTCAAG 4133
4051 ATCTGGCAAAAGATTCACAG 4110
4134 TTTGACTCAAG 4193

Db 823 CGTGGATTAATATGGAACGATTTTAAGCGCTCTTTGAATCTCAAGGAGGTAAAGGT 882
Qy 897 TCGTGGCGGACGACCTTGACCCAGCTCAAGCCAGACCCAGAGCAGAAACAAGGAAATCTGA 956
Db 883 TGGTGGCGGACGACCTTCAAGCAGCTGTGCGCAGACGACGAGCAGAAACAAGAGATCAGA 942
Qy 957 ACCCCATTTCTGAGCGCGCTGTCACTCGGCTCAAAAACCTCCGCGCGCTACATGAGAGCTGG 1016
Db 943 ATCCCAATTTCTGAGCGCGCTGTCACTCGGCTCAAAAACCTTCAAGCGGTACATGAGAGCTGG 1002
Qy 1017 TCGGCGTGTGTGTGAGCCGCGGACATCACTTCCAGAAACAGATGATCAGAGAGCAGG 1076
Db 1003 TCGGCGTGTGTGTGAGCCGCGGACATCACTTCCAGAAACAGATGATCAGAGAGCAGG 1062
Qy 1077 CTTGTGATCATCTCTTCAAGCGCGCTTCAACTCGCGGTCCCAATCAAGGCGCTCTGG 1136
Db 1063 CTTATATCATCTCTTCAAGCGCGCTTCAACTCGCGGTCCCAATCAAGGCGCTCTGG 1122
Qy 1137 ACAATGCGGCAAGATCATGCGCTGACCAAAATCGCGCGCGACTACCTGTAGGCGCCG 1196
Db 1123 ACAATGCGGCAAGATTAATGAGCTGAGCTGATAAACCGCGCGCGACTACCTGTAGGCGCCG 1182
Qy 1197 CTCCGCGCGGACATTTAAACCAACCGCATCTACCGCATCTGTAGAGCTGAACGCTACG 1256
Db 1183 AGCCGCTGAGGACATTTCCAGCAATCGGATTTAATAATTTGGAATCAACGCGGTACG 1242
Qy 1257 AACTGCGTATGCGCGCTCGCTCTTTCTCGGCTGTGGCCAGAAAGTTGGGAAAGCGCA 1316
Db 1243 ATCCCAATTAATGCGCGCTCTCTTTCTGTGGATGGGCGACGAAATAATTTGGCAAGAGGA 1302
Qy 1317 ACACATCTGTGCTGTGTGGCGCGGACCAACGCGGCAAGAACCAATGCGCGGAAACCATG 1376
Db 1303 ACACATCTGTGCTGTGTGGCGCTGTGGCGCTGTGGCGCTGTGGCGCTGTGGCGCTGTGG 1362
Qy 1377 CCCAGCGGTGCGCTTCTTACGCGCTGTGCTCAACTGACCAATGAGAACTTTCTCTTCAATG 1436
Db 1363 CCCAGCTGTGCGCTTCTTACGCGGTGTGCTCAACTGACCAATGAGAACTTTCTCTTCAATG 1422
Qy 1437 ATTGCGTGTGCAAAATGATGATCTGTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1496
Db 1423 ACTGTGTGCAAAATGATGATGATCTGTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1482
Qy 1497 AGTCCGCGCAAGGCGCATCTCGCGCGGACAGAAAGTGTGCGGTGCAACCAAAAGTCAAGTCTG 1556
Db 1483 AGTCCGCGCAAGGCGCATCTCGCGCGGACAGAAAGTGTGCGGTGCAACCAAAAGTCAAGTCTG 1542
Qy 1557 CCGCGCGAGATGACCCCAACCCCGGTGATGCTACCTTCAACCAACCAATGTGCGCGGTGA 1616
Db 1543 CCGCGCGAGATGACCCCAACCCCGGTGATGCTACCTTCAACCAACCAATGTGCGCGGTGA 1602
Qy 1617 TTGACGCGGAAACAGACCACTTGTGAGACCAAGCGCGGTGCAAGACCGGATGTTCAAT 1676
Db 1603 TTGACGCGGAACTCAACGACCTTGTGAAACCAAGCGCGGTGCAAGACCGGATGTTCAAT 1662
Qy 1677 TTGAATCTACCCGCGCTGTGAGCATGATTTGGCAAGTGAACAAGAGAGAGTCAAG 1736
Db 1663 TTGAATCTACCCGCGCTGTGATCATGATTTGGGAAAGTCAACAAGAGAGAGTCAAG 1722
Qy 1737 AGTTCTTTCGCTGGGCGCAGGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAGAA 1796
Db 1723 ACTTTTTCGCGTGGGCAAAAGGATCACTGAGTTGAGTGGAGCATGAAATTTCTACGTCAGAA 1782
Qy 1797 AGGTGAGAGCAAAAGAGCCGCGCGCATGACCGGATGAAAGAGAGAGCCCAAGGCGG 1856
Db 1783 AGGTGAGAGCAAAAGAGCCGCGCGCATGACCGGATGAAAGAGAGAGCCCAAGGCGG 1842
Qy 1857 CTTGCGCGCTAGTGGCGGATCATGAGTGAAGAGCGGAAAGAGTCCGGTGAAGCTTTG 1916
Db 1843 TGGCGAGTCAAGTTGGCGGACATGAGTGAAGAGCGGAAAGAGTTCATCACTACG 1899
Qy 1917 CCGAGAGATCAAAACCAATGTTCTGTGACGCGGCGATGCTTCAAGTGTCTTCTCT 1976

Db 1900 CAGACAGATCAAAACCAATGTTCTGTGACGCTGGGACGATCTGATGCTGTCCCT 1959
Qy 1977 GCAAGACATGCGGAGAAATGAATCAAAATTTCAAACTTTGCTTCAAGCAGCGGACGAG 2036
Db 1960 GCAAGACATGCGGAGAAATGAATCAAAATTTCAAAATTTGCTTCAAGCAGCGGACGAG 2019
Qy 2037 ACTGTTCAAGAGTCTTCCCGCGGTGTCAAAATCTCAACCGGT---GTCAGAAAGAGA 2093
Db 2020 ACTGTTCAAGAGTCTTCC---GTCAGAAATCTCAACCGGTGTGTGTGTCAAAAGG 2076
Qy 2094 CGTATGCAAACTGTGTCCATTCATGATCTGTGTGGGCGCGCTCCGAGATGCTTGTCT 2153
Db 2077 CGTATGCAAACTGTGTCTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2133
Qy 2154 CCGCGTGCAGTGTGTCAAGTGAACCTGAGATGATGATGATGATGATGATGATGATGATGAT 2213
Db 2134 CTGCTGCGGATCTGTGTCAAGTGAACCTGAGATGATGATGATGATGATGATGATGATGAT 2193
Qy 2214 AAACAGATATGAGCTCCGATGATTCCTTCCAGATTTGGCTGAGACCAACTCTGTAG 2273
Db 2194 AAATCAAGTATGCTCCGATGATTCCTTCCAGATTTGGCTGAGACCAACTCTCTGTAG 2253
Qy 2274 GCGATTCGAGATGCTGTGGGACTTGAACCTGAGACCCGAGCCCAAGCCCAACGCA 2333
Db 2254 GGAATTAAGACAGTGTGTGAAGCTCAAACTGTGCGCCACCAACCAAGCCGAGAGCGG 2313
Qy 2334 AAGCAGAGAGAGCGCGGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2393
Db 2314 CATTAAGAGACAGAGGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2373
Qy 2394 GGAATGAGAGAGAGCGCGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2453
Db 2374 GGAATGAGAGAGAGCGCGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2433
Qy 2454 GCTTACGACGAGAGCTCAAGCGGCTGTGTGCTTCTGTGCTAAGATACCTGCGGAC 2513
Db 2434 GCTTACGAGCGGAGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2493
Qy 2514 GCGGATTCAGAGAGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2573
Db 2494 GCGGATTCAGAGAGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2553
Qy 2574 GTCTTCAAGCGCAAGAGCGGCTTCTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2633
Db 2554 GTCTTCAAGCGCAAGAGCGGCTTCTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2613
Qy 2634 ACGGCTCTGTGAAGAAAGTCCGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2693
Db 2614 ACGGCTCTGTGAAGAAAGTCCGCTGTGTGCTTCTGTGCTAAGATACCTGCGGACCTTCAAC 2673
Qy 2694 GGCATGCGGCAAGAGCGGCGGCTTCAAGAAAGTCTTCAAGATACCTGCGGACCTTCAAC 2753
Db 2674 GGCATGCGGCAAGAGCGGCGGCTTCAAGAAAGTCTTCAAGATACCTGCGGACCTTCAAC 2733
Qy 2754 GACTCAGAGTCACTCCGATTCACCAACTCTGTGAGAACTTCAAGAACTTCAAGAACTTCAAG 2813
Db 2734 GACTCAGAGTCACTCCGATTCACCAACTCTGTGAGAACTTCAAGAACTTCAAGAACTTCAAG 2793
Qy 2814 GTGGAGCTTACTTCAATGCTTGTGAGCGGTGTGAGCGGACCAATGCGGACCAATTAAGAGG 2873
Db 2794 GTGGAGCTTACTTCAATGCTTGTGAGCGGTGTGAGCGGACCAATGCGGACCAATTAAGAGG 2853
Qy 2874 GCGGAGCGAGTGGGATGCTTCAAGAAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAAT 2933
Db 2854 GCGGAGCGAGTGGGATGCTTCAAGAAATTTGCAATTTGCAATTTGCAATTTGCAATTTGCAAT 2913
Qy 2934 AAGATCATCAACCAAGACCCGACCTGAGCCCTTCCCACTTCAATTAACCACTTCAAC 2993
Db 2914 AAGATCATCAACCAAGACCCGACCTGAGCCCTTCCCACTTCAATTAACCACTTCAAC 2973
Qy 2994 AAGCAATCTCCAGTGTCTTCAAGCGGCGGACGAGCAAGCAACCACTTCTGCTACAG 3053
Db 2974 AAGCAATCTCCAGTGTCTTCAAGCGGCGGACGAGCAAGCAACCACTTCTGCTACAG 3030

OY	3054	ACCCTTGGGGGATTTTGGATTTTGAACAGATTCCATGCGCACTTTTACACAGGACCTGG	3113
Db	3051	ACCCTTGGGGGATTTTGGATTTTGAACAGATTCCATGCGCACTTTTACACAGGACCTGG	3090
OY	3114	CAGCGACTCATCAACAACAATTTGGGGAATTCGGGCCCAAGAGACTCAACTCAACTCTTC	3173
Db	3091	CAAGAGCTCATCAACAACAATTTGGGGAATTCGGGCCCAAGAGACTCAACTCAACTCTTC	3150
OY	3174	AACATTCGAAGTCAAGAGGTGACGAGCAATGATGCGGTACACACCATGCGTAAATPACTT	3233
Db	3151	AACATTCGAAGTCAAGAGGTGACGAGCAATGATGCGGTACACACCATGCGTAAATPACTT	3210
OY	3234	ACCGACAAGTTCAGTCTTCTCGGACTCGAGTACCAAGCTTCGTAAGTCTCGGCTCT	3293
Db	3211	ACCGACAAGTTCAGTCTTCTCGGACTCGAGTACCAAGCTTCGTAAGTCTCGGCTCT	3270
OY	3294	GGCGACCAAGGGGTGCTCTCCCTCCCTCCGTTCCCGGGAGCGTGTTCATGATTCGGCAATAGCG	3353
Db	3271	GGCGATCAAGGATGCTCTCCCGCGGTTCCAGAGAGCGTCTTCAATGCTGCCACAGATAGGA	3330
OY	3354	TACCTGACGCTCAACAATGCGAGCGCAACCGGGGACGTTATCATCTTTACGTGCGTGGAA	3413
Db	3331	TACCTGACGCTCAACAATGCGAGCGCAACCGGGGACGTTATCATCTTTACGTGCGTGGAA	3390
OY	3414	TATTTCCCTTCTCAGATCTGAGAACGGGCAACAATTACCTTCACTACACTTTGAG	3473
Db	3391	TACTTTCTTCTCAGATCTGCGTACCGGAAACAATTACTTCACTACACTTTGAG	3450
OY	3474	GAAAGCGCTTTCACAGAGCTACGCGGACAGCCAGACCTCGAGCCGCTGATGAATCT	3533
Db	3451	GACGTTCTTTCACAGAGCTACGCGTACAGCCAGATCTGAGCCGCTGATGAATCT	3510
OY	3534	CTCATGCAACATACCTGTATTTACCTGTAACAGAACTCAAAATCACTCGGAGAGTGCCTAA	3593
Db	3511	CTCATGCAACATACCTGTATTTACCTGTAACAGAACTCAAAATCACTCGGAGAGTGCCTAA	3570
OY	3594	AACAAGAGACTGTGCTGTTTAAAGCCGTGGGTCTCAAGCTGCAATGTCTGTTACGCCAATAAC	3653
Db	3571	CAGTCMAAGCTTCAGTTTTCTCAGGCCCGAGAGCGAGTGCATTTCCGGGACCACTCTAGAGAC	3630
OY	3654	TGGCTACCTGAGACCCTGTTTATTCGGACAGAGCGGCTTTTAAACAAAAAGACAAACAAC	3713
Db	3631	TGGCTACCTGAGACCCTGTTTATTCGGACAGAGCGGCTTTTAAACAAAAAGACAAACAAC	3690
OY	3714	AACAGCAATTTTAACTGAGCTGTGTCTTCAAAATATTAACCTCAATGGCGGTGAATCCATC	3773
Db	3691	AACAGTGAATCTGCTGAGCTGTGTCTTCAAAATATTAACCTCAATGGCGAGACTCTCTG	3750
OY	3774	ATCAACCTGTGCACTGTATATGGCTCTCAACAAGATCAAGACCAAGTTCTTTCCATG	3833
Db	3751	GTCGAATCGGGGCGCCGCGCATGCAAGCCCAAGAAGCAATGAAGAAAGTTTTTCTCAG	3810
OY	3834	AGCGGTGTCATGATTTTTTGGAAAAAGAGACGCCCGGAGCTTCAAACTGCAATTGGAACAT	3893
Db	3811	AGCGGGGTTCATCTTTTGGGAAGCAAGGCTCAAGAAAAAATAATGTGAATGAAAG	3870
OY	3894	GTCATGATTTACAGACGAAGAGAAATTAAGCCACTAACTCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTTACAGACGAAGAGAAATCGAAACAACATCCCGTGGCTTACCGAACAGATAT	3930
OY	3954	GGAGACCGTGGAGCAATTTCCAGAGACGACGACAGCCCTGTGGACCGGAGATGTGCAT	4013
Db	3931	GGTTCTGTATTTACCAACCTCCAGAGAGGCAACAGACAGCTTACCGCAAGATCAAC	3990
OY	4014	GCTATGGAGCAATTAACCTGGCATGGTGTGGCAAGATAGACGTTACTTGCAGAGGTCC	4073
Db	3991	ACACAAGCGTTCTTCCAGCGATGGTCTGGCAGACAGAGATGTACTTCAAGGGGCC	4050
OY	4074	ATTGGGCGCAAAATTTCTCAACAAGATGGAACATTTACCCGTCTCTTATATGGGCGGC	4133
Db	4051	ATCTGGGCAAAAGATTTCAACAAGACGGAACATTTTACCCCTCTCTCATATGGGTGGA	4110

QY	4134	TTTGGACCTCAAGAAACCCGGCTCTCAATCTCTATCAAAAAACAGCGTGTCTCGCAAT	4133
Db	4111	TTTCGACTTAAACACCTCTCTCCACAGATTTCTCAAGAAACCCGGTACTCGCAAT	4170
QY	4194	CCTCCGGGAGTTTTAGACTCAAGATTTTGCTTCATTCATCACCAAATCTCCACAGGA	4253
Db	4171	CCTTCGACCACTTCAAGGGGGGAAAGTTTGCTTCTTCATCAACAAGTACTCACGGGA	4230
QY	4254	CA-AGTAGTGTGAAATTGAATGGAGCTGCAGAAAAGAAAACGCAAGCGCTGGAATCC	4312
Db	4221	CACGGTCAGCGTGAAGATCGAGTGGAGCTGCAGAAAGAAAACGCAAAACGCTTGAAATCC	4290
QY	4313	CGAAGTGAGTACATCAATCAATTATGCAAAATCGCAACGGTATTTTACTGTGCAAA	4372
Db	4281	CGAAATTCAGTACCTTCCACTCAACAAAGCTGTTAAATCTGTGACTTACCGTGAATAC	4350
QY	4373	CAATGCACTTTATACTGAGCTCGCCCATTTGGACCCCGTACCTTACCCGCTCCGTGA	4432
Db	4351	TAAATGGCGTATTCAGAGCGCTCGCCCATTTGGACCAAGATCTGATCTGTA	4410
QY	4433	ATTACGCTTATCAATTAACCGGTTGATTTGTTCACTTGAACTTTGTCTCTGTCT	4492
Db	4411	ATTCTGTTTATCAATAAACCGTTTAATTCGTTCAGTTGAACTTTGTGTTCTCGCTAT	4470
QY	4493	TCTTATCTEATCGGTTTACAAATGTTTAAAGTTACATCAATTAACGCTTGTGGGCTTCGC	4552
Db	4471	TTCTTTCTTATCTAGTTTTCATGCTCAAGTACGTAAATTAATTAAGATGGCGGTTAATCATTTAA	4530
QY	4553	GATTAAGAACTTACGTATCGGTTTACCCCTAGTGAATGATGGTCCCACTTCCCTCTGC	4612
Db	4531	CTACACAGGA-----AACCTAGTGAATGGAAGTTGGCACTTCCTCTCTGC	4574
QY	4613	GGCGTCGCTCGCTCGGTGGGGGCTGCAGAACCAAAAGTCCGAGACGGCAAGCTCTGCTC	4672
Db	4575	GGCGTCGCTCGCTCACTGAAGGCGAGGACCAAAAGTCTGCGCCGACGCGCCGAGCTTTGCCC	4634
QY	4673	TGCGGGGCCCCAGAGGAGCGAGCGGCGAGAGAGGGAGTGGGCAA	4718
Db	4635	GGGCGGCTCAGTAGGAGGAGGACCGGCAAGAGGGGAGTGGCCAA	4680

RESULT 6
 US-10-263-127-1
 Sequence 1, Application US/10263127
 Publication No. US20030082145A1
 GENERAL INFORMATION:
 APPLICANT: Johnson, Philip R.
 TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 6300 Sears Tower, 233 S. Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/263,127
 FILING DATE: 02-Oct-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/10/077,294
 FILING DATE: 15-Feb-2002
 APPLICATION NUMBER: 09/691,604
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: No. US20030082145A1and, Greta E.

REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-263-127-1

Query Match 65.3%; Score 3082.4; DB 15; Length 4680;
Best Local Similarity 80.0%; Pred. No. 0;
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCGTGGGGCTGCGCAACCAAGTCCG 60
DB 1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCGTGGGGCTGCGCAACCAAGTCCG 60
QY 61 AGACGGAGAGCTCTGCTGCGCGCCCAACCGAGCGAGCGCGAGAGAGAGT 120
DB 61 AGACGGCGCGGCTTTGGCCCGGGCGCTCACTGAGCGAGCGCGCGAGAGAGAGT 120
QY 121 GGCACCTCATCTAGGGGTAATCGGAGCGCTCCACGCTGCGCGCTGACGCTGA 180
DB 121 GGCACCTCATCTAGGGGTAATCGGAGCGCTCCACGCTGCGCGCTGACGCTGA 180
QY 181 CGTAATTAAGTCATAGGG---GAGTGGTCTGTATTAGCTATGACGAGTGTCTTTC 237
DB 164 CGTAATTAAGTCATAGGGTTAGAGAGTCTGTATTAGAGTCACTGAGTGT---TTTTC 222
QY 238 GACATTTTTCGACACCACTGCGCATTTTATATATGCGAGTGAAGGACAGAT 297
DB 223 GACATTTTTCGACACCACTGCGCATTTTATATATGCGAGTGAAGGACAGAGT 282
QY 298 CTCATTTTGG-AACGCGAATTTTGAAGAGACGACGCGATGCGGGCTTTCAAGATCG 356
DB 283 CTCATTTTGAAGGCGGAGGTTTGAAGCGCGACGCGCATGCGGGGTTTACCAAGATG 342
QY 357 TGAATCAAGGTGCGAGGACCTGACGAGCACTGCGGGCATTTTGAAGTCTGTTTGA 416
DB 343 TGAATTAAGTCCCAAGACTTGAAGGCACTGCGCGCATTTTGAAGTCTGTTTGA 402
QY 417 GCTGGTGGCGAGAGAGATGAGAGTCCCGGATTTCTGACATGATCTGAATCTGA 476
DB 403 ACTGGTGGCGAGAGAGATGAGAGTGGCGCGCATTTCTGACATGATCTGAATCTGA 462
QY 477 TTGAGCAGGCAACCCCTGACCGTGGCGAGAGCTGACGCGCATTTCTGTGTCATAGGC 536
DB 463 TTGAGCAGGCAACCCCTGACCGTGGCGAGAGCTGACGCGCATTTCTGACGAGATGCG 522
QY 537 GCGCGTGAAGAGAGAGAGGCGCTCTTTTGTTCAGTTGAGAGAGGAGGAGTCT 596
DB 523 GCGCGTGAAGAGAGAGAGGCGCTTTTGTTCAGTTGAGAGAGGAGAGTCT 582
QY 597 ACTTCACTTCATATTTCTGTGAGACCAAGGAGGTCATTCATGATGCTGAGCGCT 656
DB 583 ACTTCACTATGCACTGCTGTGAGAACCAAGGAGGTCATTCATGATGCTTGGAGGTT 642
QY 657 TCTGAGTCAAGATTAAGGACAGCTGTCAGACCATCTACGCGGATGAGCGCACCC 716
DB 643 TCTGAGTCAAGATTAAGGACAGCTGTCAGACCATCTACGCGGATGAGCGCACCT 702
QY 717 TGCCCACTGGTGGCGGTGACCAAGACGCGTAATGGCGCGAGGGGAGAACAAAGTGG 776
DB 703 TGCCCACTGGTGGCGGTGACCAAGACCAAGATGGCGCGAGGGGAGAACAAAGTGG 762
QY 777 TGACGAGTGTACATCCCAACTACTCTGCGCCAAAGACTCAAGCCGAGCTGACGTGG 836

DB 763 TGATAGAGTGTACATCCCAATTACTGTCTCCCAAAACCAGCTGAGCTCAGTGGG 822
QY 837 CGTGAATTAACATGAGAGAGTATATAGGCGCTTTTGAACCTGGCGGCGCAACGCG 896
DB 823 CGTGAATTAATAGAAAGATATTTAAGCGCTTTTGAATCTACAGAGCGCTAAGCGGT 882
QY 897 TCGTGGCGAGCACTGACCCAGTCAAGAGAGAGAGAGAGAGAGAGAGATCTGA 956
DB 883 TCGTGGCGAGCACTGACCCAGTCAAGAGAGAGAGAGAGAGAGAGAGATCTGA 942
QY 957 ACCCAATTTGACGCGCTGTTCATCCGCTCAAAAACCTTCGCGCGTACATGAGCTGG 1016
DB 943 ATCCCAATTTGATGCGCGCTGTTCATCAAAAACCTTCAGCGAGTACATGAGCTGG 1002
QY 1017 TCGGCTGGCTGTGAGACCGGGGATCATCCTCGAGAGAGAGAGATCCAGAGACAGG 1076
DB 1003 TCGGCTGGCTGTGAGACAGGGGATTAATCTCGAGAGAGAGAGATCCAGAGACAGG 1062
QY 1077 CTTGTATCATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGGCGCTCTGG 1136
DB 1063 CTTGTATCATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGGCGCTCTGG 1122
QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAATTCGCGCTCACTACCTGTAGGCGCG 1196
DB 1123 ACAATGCGGCAAGATCATGCGCTGACCAATTCGCGCTCACTACCTGTAGGCGCG 1182
QY 1197 CTCGCGCGCGGCAATTAACCAACCGCATCTACCGGATCCGAGGCTGAAAGGCTACG 1256
DB 1183 AGCCCGTGAAGAGATTTCCAGATTCGAGATTTATTAATTTTGAATTAAGGAGTACG 1242
QY 1257 AACCTGCTATGCGCGGCTCCGCTTTTCTCGGCTGGGCGCGCAAAAGGTTGGAGACGA 1316
DB 1243 ATCCCAATTAATGCGGCTCCGCTTTCTGAGAGAGGCGCAAGAAAAGTTGGAGAGGA 1302
QY 1317 ACAACATCTGGCTGTTTGGGCGCGGCAACGCGGCAAGCAATTCGCGGAAAGCCATCG 1376
DB 1303 ACAACATCTGGCTGTTTGGGCGCTGCAACTACCGGAGAACCAATTCGCGGAAAGCCATCG 1362
QY 1377 CCGACGCGGCGCTCTACGCGTGTGCAATGAGCAATGAGAACTTTCCCTTCAATG 1436
DB 1363 CCGACGCTGCGCTCTTACGCGTGTGCAATGAGCAATGAGAACTTTCCCTTCAACG 1422
QY 1437 ATTGCGTCAAGATGATGATGTTGGAGAGAGAGGCAAGATGAGCGCCAAAGTCTGG 1496
DB 1423 ACTGTGTCAAGATGATGATGTTGGAGAGAGAGGCAAGATGAGCGCCAAAGTCTGG 1482
QY 1497 AGTCCGCAAGGCTATTTCTGCGCGGCAAGAGTGGCGCTGACCAAAAGTCAAGTCTG 1556
DB 1483 AGTCCGCAAGGCTATTTCTGCGCGGCAAGAGTGGCGCTGACCAAAAGTCAAGTCTG 1542
QY 1557 CCGCGCAAGTCAAGGCTCAAGCGCGGATGATGATGATGATGATGATGATGATGATGATG 1616
DB 1543 CCGCGCAAGTCAAGGCTCAAGCGCGGATGATGATGATGATGATGATGATGATGATGATG 1602
QY 1617 TTGACGGAAACAGACCACTTTCAGACACAGAGCGTTTTCAGAGACCGGATTTCAAT 1676
DB 1603 TTGACGGAAACAGACCACTTTCAGACACAGAGCGTTTTCAGAGACCGGATTTCAAT 1662
QY 1677 TTGAATCAACCGCGCTCTGAGACATGCTTTGGCAAGGTGACAAAGCAGAAAGTCAAG 1736
DB 1663 TTGAATCAACCGCGCTCTGAGACATGCTTTGGCAAGGTGACAAAGCAGAAAGTCAAG 1722
QY 1737 AGTTCTTCGCGTGGGAGAGATCAAGTGAAGAGGAGGAGTGAAGTTTACGTCAAG 1796
DB 1723 AGTTCTTCGCGTGGGAGAGATCAAGTGAAGAGGAGGAGTGAAGTTTACGTCAAG 1782
QY 1797 AGGATGAGCAACAAAGAGACCGCGCGATGACCGGATTAAGAGAGAGAGAGAGAGAG 1856
DB 1783 AGGATGAGCAACAAAGAGACCGCGCGATGACCGGATTAAGAGAGAGAGAGAGAGAG 1842
QY 1857 CTTGCCCTTCAAGTGGCGGATTCATGACGTCAAGCGGAGAGAGAGTCTGGTGAAGCTTTG 1916

1843 TGGCGAGTCAGTTGCGGAGCATCGACGTGACGCGGA---AGTTGATCACTACG 1899
QY 1917 CCGACAGTACCAAAATAATGTTCTGTACGCGGGCATGCTTCAAGTGTCTTCCCT 1976
DB 1900 CAGACAGTACCAAAATAATGTTCTGTACGCGGGCATGAACTGTATCTGTCTCCCT 1959
QY 1977 GCAAGACATCGAGAGATGATGATCAGAAATTTCAATTGCTTCAAGCGAGAG 2036
DB 1960 GCAGACATCGAGAGATGATGATCAGAAATTTCAATTGCTTCAAGCGAGAG 2019
QY 2037 ACTGTTCAAGTGTCTTCCCGCGGTGTGAGATCTCAACCGGT---CGTCAAGAAAGA 2093
DB 2020 ACTGTTAGAGTGTCTTCC---CGTGTCAAGATCTCAACCGGTCTGTGTCTCAAAAGG 2076
QY 2094 CGATTCGAAATCTGTGCGATTCATCATCTGTGCGGCGGGCTCCGAGATGTCTTCT 2153
DB 2077 CGATTCGAAATCTGTGCGATTCATCATCTGTGCGGAGAAAGGTGCCGA---CGCTTGCA 2133
QY 2154 CCGGCTGCGATTCGTGCAAGTGTGACCTGTGATCTGTTTCTGAGCAATTAATGACTT 2213
DB 2134 CTGCTGCGATTCGTGCAATGTGATTTGATGATCTGATCTTTGAACTTAATTAATTT 2193
QY 2214 AAACAGATATGCTGCGATGATTTATCTTCCAGATTTGCTGAGCAAACTTCTGAG 2273
DB 2194 AAATCAGGTATGCTGCGATGATTTATCTTCCAGATTTGCTGAGCACTCTCTGAA 2253
QY 2274 GGCATTCGCGATGTGTGGACTTGAAACTGTGAGCCCGAAAGCCCAAGCCACAGCA 2333
DB 2254 GGAATATGACATGTGTGGAAGCTCAACCTGCGCCACACACCAAGCCCGAGAGCGG 2313
QY 2334 AAGCAGAGCAAGCGCGGGGTGTGTGCTTCTGTGCTACAAAGTACCTGAGACCTTTCAAC 2393
DB 2314 CATAGAGCAACAGCAGAGGGGTCTTGTGTCTTCTGTGATCAAGTACCTGAGACCTTTCAAC 2373
QY 2394 GAACTGCAAGAGGAGAGCCGTCAACGCGCGGACCGAGCCCTGAGCAAGCAAG 2453
DB 2374 GAACTGCAAGAGGAGAGCCGTCAACGAGGCAAGACCGCGGCTGAGCAAGCAAA 2433
QY 2454 GCTTACGACCAAGCTTCAAAAGCGGTGACATTCCTGATCTGTGATTAACCAAGCGGAC 2513
DB 2434 GCTTACGACCAAGCTTCAAAAGCGGTGACATTCCTGATCTGTGATTAACCAAGCGGAC 2493
QY 2514 GCGGAGTTTCAAGAGCGGTGCAAGATACGTTTGGGGCAACTCGGCGAGCA 2573
DB 2494 GCGGAGTTTCAAGAGCGGTGCAAGATACGTTTGGGGCAACTCGGCGAGCA 2553
QY 2574 GTCTTCAGGCGCAAGAGCGGTGCTGAACTTCTGCTGTGTTGAGAAAGCGCTAAG 2633
DB 2554 GTCTTCAGGCGCAAGAGCGGTGCTGAACTTCTGCTGTGTTGAGAAAGCGCTAAG 2613
QY 2634 ACGGCTCTGGAAGAAAGCTCGGTGAGAGCTGCGCAAGAGCCAGACTCTCTCTCG 2693
DB 2614 ACGGCTCTGGAAGAAAGAGCGGTGAGAGCTGCTGTGAGCGAGACTCTCTCTCG 2673
QY 2694 GGCATGCGGCAAGCAGGCGCGAGCGGCTAAAAAGAGCTCAATTTTGTGAGACTGCG 2753
DB 2674 GGAACCGGAAGAGCGGCGCGAGCGCTCAAGAAAAAGATTTTGTGAGACTGCG 2733
QY 2754 GACTCAGAGTCAGTCCCGATCCCACTCTCGAGAACTCTCAGCAACCCCGCTGCT 2813
DB 2734 GAGCGAGACTCAAGTACTGACCCCACTCTCGAGCAAGCAACGAGCGCCCTCTGCT 2793
QY 2814 GTGAGACTTCTACATAGCTTCAAGCGGTGCGGCACTCAATGCGACATTAAGAGG 2873
DB 2794 CTGGGAACTAATAGATGCTACAGGAGTGGCGCACTAATGCGACATTAAGAGG 2853
QY 2874 GCGGAGGAGTGGGTATGCTGAGAAATTTGAGATTTGAGATTTGAGAGTGGGAG 2933
DB 2854 GCGGAGGAGTGGGTATTTCTCTCGAAATTTGGATTTGAGATTTGAGAGTGGGAG 2913
QY 2934 AGAGTCACTACCAACAGACCCGAGCTGCGGCTGTGCGCACTTAACAATAACCTCTAC 2993
DB 2914 AGAGTCACTACCAACAGACCCGAGCTGCGGCTGTGCGCACTTAACAACAACCTCTAC 2973

2994 AAGCAATCTCCAGTGTCTCAACGCGGCGCAGCAAGCAACCACTTCTGGCTACAG 3053
QY 2974 AAACAAATTTCCAGCAATCA---GGAGCTCGAAGCAATCACTTCTGGCTACAG 3030
DB 3054 ACCCGTGGGGGATTTTGAATTTCAACAGATTCCTGCACTTTTCAACAGCTGAG 3113
QY 3031 ACCCTTGGGGGATTTTGAATTTCAACAGATTTCACTGCACTTTTCAACAGCTGAG 3090
DB 3114 CAGGACTCATCAACCAATTTGGGATTTCCGCGCAAGAGACTCAACTTCAACTCTTC 3173
QY 3091 CAAGACTCATCAACCAACTGGGGATTTCCGACCAAGAGACTCAACTTCAACTCTTC 3150
DB 3174 AATATCAAGTCAAGAGGATCAAGATGATGAGGGGTCAACCACTGCTAATTAACCTT 3233
QY 3151 AATATCAAGTCAAGAGGATCAAGATGATGAGGGGTCAACCACTGCTAATTAACCTT 3210
DB 3234 ACCAGACGGTTCAATCTTCTGAGCTGAGATCAAGCTTCCGATACCTCGGCTCT 3293
QY 3211 ACCAGACGGTTCAAGTGTTTTACTGACTGAGATCAAGCTTCCGATACCTCGGCTCT 3270
DB 3294 GCGCACAGGGCTGCTCTCTCTCTGTTCCGCGGAGCGTGTATGATTCGCAATTCGG 3353
QY 3271 GCGCATCAAGAGATGCTCCCGCGCTTCCAGCAGAGCTCTCATGTGTGCAAGATGAG 3330
DB 3354 TACCTGAGCTCAACATGAGCAGCCGCTGGAGCTTCACTTTTACTGCTGGA 3413
QY 3331 TACCTGAGCTCAACATGAGGAGTCAAGATGAGGAGCTTCACTTTTACTGCTGGA 3390
DB 3414 TATTTCTCTCTGAGATGCTGAGAGCGGCAACCTTACTTCACTTCACTTCTTGA 3473
QY 3391 TACTTCTCTCTGAGATGCTGAGTACCGAAGCACTTACTTCACTTCACTTCTTGA 3450
DB 3474 GAAATGCTTTCACAGAGCTACGCGCAACGCAAGCCTGAGCCGCTGATGAATCT 3533
QY 3451 GAAATGCTTTCACAGAGCTACGCGCAACGCAAGCCTGAGCCGCTGATGAATCT 3510
DB 3534 CTGATGACCAATCACTGATTTTACCTGAACAGAACTCAAAATGATCCGGAATGCGCAA 3593
QY 3511 CTGATGACCAATCACTGATTTTACCTGAACAGAACTCAAAATGATCCGGAATGCGCAA 3570
DB 3594 AACAGAGACTGTGTTTATGAGCGGTGCTCCAGCTGCAATGCTGTGACGCCCAAAAC 3653
QY 3571 CAGTCAAGCTTCAAGTTTCTCAAGCGCGAGCGAGCAATTTGGGACCACTTGAAGAC 3630
DB 3654 TGGCTACTGAGACCTGTTATTCGAGCAGCGCGTCTTAAACCAAAACAGCAACAC 3713
QY 3631 TGGCTACTGAGACCTGTTATTCGAGCAGCGCGATTCAAAGCATCTGCGATTAACAC 3690
DB 3714 AACAGCAATTTTACTGAGTGTGCTTCAAAATTAATTAATCAATGAGGCTGATCATC 3773
QY 3691 AACAGCAATTTTACTGAGTGTGCTTCAAAATTAATTAATCAATGAGGCTGATCATC 3750
DB 3774 ATCAACCTGAGCACTGATTTGAGCTTCAAAAGAGCAAGCAAGTCTTCTTCCATG 3833
QY 3751 GTGAATCCGAGCGCGCGCATGCAAGCAACAGAGCAATGAAGAAATTTTCTCTCAG 3810
DB 3834 AGCGGTGTCAATTTTGGAAAAAGAGCGCGAGCTTCAACACTGATTTGACAT 3893
QY 3811 AGCGGGGTCTCAATTTTGGGAAAGCAAGGCTCAGAAAAAATGTAATTTGAAG 3870
DB 3894 GTGATGATTTCAAGAGAGAAATTAAGCACTTACCTGTGTGCGACCGAAAGTTT 3953
QY 3871 GTGATGATTTCAAGAGAGAAATTAAGCACTTACCTGTGTGCGACCGAGTAT 3930
DB 3954 GGGACGTGCAATTTTCCAGAGCAGAGCAAGCCCTGCAAGCGGAGTGTGAT 4013
QY 3931 GGTCTGTATCTTCACTTCCAGAGAGGCAAGCAAGCACTTACCGAGATGTCAAC 3990
DB 4014 GTATGAGGCAATTTACCTGCAATGTGTGCAAGATGAGACGTGTACCTGAGGCTCC 4073
QY 3991 ACACAAGGCGTCTTCCAGGCAATGTGTGCAAGATGAGACGTGTACCTGAGGCGCC 4050

Db 1063 CCTCATCATCTCTCTCATATGCGGCTCCAACTCGCGGTCCCAATCAAGGCTGCTTGG 1122
Qy 1137 ACAATGCGGCAAGATATGCGCTGACCAATTCGCGCCGACTACTGTAGGCCCCG 1196
Db 1123 ACAATGCGGCAAGATATGAGCTGCTAAACCGGCCGACTACTGTAGGCCCCG 1182
Qy 1197 CTCGCGCGGCAATTAACCAACCGCATCTACCGCATCTGAGCTGAACGCGTACG 1256
Db 1183 AGCCCGGAGGACATTTCCAGCATGTGATTTAAATTTTGAATTAACGCGTACG 1242
Qy 1257 AACCTGCTAAGCGCGCTGCTTCTCGCTGCGCCGCAAAAAGTTGCGGAGCGCA 1316
Db 1243 ATCCCAATATGCGGCTTCCGTCTTCTGGAATGCGCCGCAAAAAGTTGCGGAGCGCA 1302
Qy 1317 ACACCATTTGGCTTTTGGGCGCGCCACACGCGGCAACCAATCTGCGGAAGCCATCG 1376
Db 1303 ACACCATTTGGCTTTTGGGCGCGCCACACGCGGCAACCAATCTGCGGAAGCCATCG 1362
Qy 1377 CCCACGCGGCTGCTTCTACGCGCTGCTCAACTGACCAATGAGAACTTCCCTCATG 1436
Db 1363 CCCACGCTGCTTCTACGCGGCTGCTTAATCTGACCAATGAGAACTTCCCTCATG 1422
Qy 1437 ATTGCGTCAACAAATGATGTGATGTGAGAGAGGCAAGATGACGCGCAAGCTGCTG 1496
Db 1423 ACTGTGCAACAAGATGTGATGTGATGTGAGAGAGGCAAGATGACGCGCAAGCTGCTG 1482
Qy 1497 AGTCCCGCAAGGCTTCTGCGCGCGCGCAAGGTCGCGTGAACCAAAATGCAAGTCT 1556
Db 1483 AGTCCCGCAAGGCTTCTGAGAGAGAGCAAGGTCGCGTGAACCAAAATGCAAGTCT 1542
Qy 1557 CGCCCGCATGACCCCAACCCCGCTGATGCTCACTCAACCAACATGTGCGCGTGA 1616
Db 1543 CGCCCGCATGAGACCCGATCTCCGTGATGCTCACTCAACCAACATGTGCGCGTGA 1602
Qy 1617 TTGACGAGAAACAGACCACTTCAAGACCAAGCCGTTGACAGACCGGATGTTCAAT 1676
Db 1603 TTGACGAGAACTCAACGACTTCAAGACCAAGCCGTTGACAGACCGGATGTTCAAT 1662
Qy 1677 TTGAACTCAACCCCGCTTGTGAGATGATCTTTGGCAAGTGAACAAAGAGTCAAG 1736
Db 1663 TTGAACTCAACCCCGCTTGTGATGATCTTTGGCAAGTGAACCAAGAGTCAAG 1722
Qy 1737 AGTTCTTCCGCTGAGCGCAGATCAAGTGAACGAGTGGCGCATGATGTTCACTGAA 1796
Db 1723 ACTTTTCCGCTGAGCGCAAGATCAAGTGTGAGTGAAGCATGATTTTCACTGAA 1782
Qy 1797 AGGATGAGCAACAAAGACCCGCGCCGATGACGCGGATTAAGCGAGCCCAAGCGG 1856
Db 1783 AGGATGAGCAACAAAGACCCGCGCCGATGACGCGGATTAAGTGAAGCCCAAGCGG 1842
Qy 1857 CCGCCCTTCAAGTGGCGGATTCATCAAGTCAAGCGGAGAGAGCTCCGCTGAGCTTGG 1916
Db 1843 TCGCGAGTCAAGTGGCGGACCATCAAGTCAAGCGGAGAGAGCTCCGCTGAGCTTGG 1899
Qy 1917 CGCAGAGTCAACAAATGTTCTGTCAGCGGCGCATGCTTCAAGTCTGTTCCCT 1976
Db 1900 CAGACAGTCAACAAATGTTCTGTCAGCGGCGCATGATCTGATCTGTTCCCT 1959
Qy 1977 GCAAGCATGCGAGAGATGAATCAGAAATTTCAATTTGCTTCAAGCGAGCGAGAG 2036
Db 1960 GCAAGCATGCGAGAGATGAATCAGAAATTTCAATTTGCTTCAAGCGAGAGAG 2019
Qy 2037 ACTGTTCAAGTGTCTTCCCGCGGTGTCAGAAATCTCAACCGGT---CGTCAAGAAAGGA 2093
Db 2020 ACTGTTCAAGTGTCTTCC---CGTCAAGAAATCTCAACCGGTCTGTCGTCAGAAAGG 2076
Qy 2094 CGATCGGAACCTGTCGCTTCAATCATCTGCGGCGCGGCTCCCGGATGTTGCTTGT 2153
Db 2077 CGATCGGAACCTGTCGCTTCAATCATCTGCGGAGAGAGGTCGAG---CGCTTGA 2133
Qy 2154 CGGCTGCGATCTGTCAGAGTGAACCTGAGTGAATGCTGTTGTCGCAATTAATGACTT 2213
Db 2134 CTGCTGCGATCTGTCAGATGGAATTTGATGATGATCTTCTTGAACATTAATGATTT 2193

Qy 2214 AAACAGGTAATGCTGCGGATGTTATCTTCCAGATTTGCTGAGGACAAACCTCTGAG 2273
Db 2194 AAATCAAGTAATGCTGCGGATGTTATCTTCCAGATTTGCTGAGGACAACTCTCTGAA 2253
Qy 2274 GGCATTGCGAGTGTGAGGACTTGAACCTGAGCCCGCAAGCCCAAGCAACAGCA 2333
Db 2254 GGAATTAAGACAGTGTGAGGACTTGAACCTGAGCCCGCAACCAAGCCCGAGAGCGG 2313
Qy 2334 AAGCAGACGACGCGCGGAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2393
Db 2314 CATAGAGACGACGAGGAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2373
Qy 2394 GGACTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2452
Db 2374 GGACTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2433
Qy 2453 GGCCTACGACAGACGCTCAAAAGCGGAGTGAACATCTGCTGCTGCTGCTGCTGCTGCT 2512
Db 2434 AGCTTACGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2493
Qy 2513 CGCGAGTTTCAAGAGGCTGCTGCAAGAGATGCTTTTGGGAGCACTTGGGCGAGC 2572
Db 2494 CGCGAGTTTCAAGAGGCTGCTTAAAGAGATGCTTTTGGGAGCACTTGGGCGAGC 2553
Qy 2573 AGTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2632
Db 2554 AGTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2613
Qy 2633 GAGGCTCTCTGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2692
Db 2614 GAGGCTCTCTGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2673
Qy 2693 GGGCATGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2752
Db 2674 GGGAGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2733
Qy 2753 GGACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2812
Db 2734 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2793
Qy 2813 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2872
Db 2794 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2853
Qy 2873 CGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2932
Db 2854 CGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2913
Qy 2933 CAGAGTCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2992
Db 2914 CAGAGTCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2973
Qy 2993 CAAGCAATCTCCAGTGTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3052
Db 2974 CAAGCAATCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3030
Qy 3053 CACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3112
Db 3031 CACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3090
Qy 3113 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3172
Db 3091 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3150
Qy 3173 CAACATCAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3232
Db 3151 TAAATTTCAAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3210
Qy 3233 TACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3292
Db 3211 TACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3270

QY	3293	TGCGCACACGAGGGCTGCCTCCCTCGTTTCCCGCGGACGTGTTCAATGATTCGCAATACGG	3352
Db	3271	GGCGCATCAAGGATGCTCCCGCGCTTCCACAGACACTCTTCATGGTGGCCACAGTAAG	3330
QY	3353	CTACCTGACGCTCAACAAATGGACGAAAGCCGTGGGACGTTCACTCTTTAATGCTGGA	3412
Db	3331	ATACCTCACTCTGAAACAAACGGGAGTCAAGGACAGTGAAGCGCTTCATTTTACTGCTGGA	3390
QY	3413	ATATTTCCCTCTCAGATGCTGAGAAAGGGGAAACAATTACCTTCAGCTACACTTGA	3472
Db	3391	GTACTTTCCTTCTCAGATGCTGCGTACCGGAAACAATTACCTTCAGCTACACTTGA	3450
QY	3473	GGAATGCTTTTCCACAGACACTACGCGACAGCCAGACGCTGGAACGGCTGATGATCC	3532
Db	3451	GGAAGTTCCTTCCACAGACACTACGCTACAGCGACAGTCTGGACCGCTCATGAAATCC	3510
QY	3533	TCTCATGACCAATTAACCTGTATTAACCTGAAACAGAACTCAAAATCAATCCGGAAGGCCA	3592
Db	3511	TCTCATGACCAATTAACCTGTATTAACCTGAAACAAACTCCAAAGTGAACCAACAC	3570
QY	3593	AAACAAGACATGCTGTTTAAACCGGTGCTCCAGCTGGCAATGTCGTTTCAACCCCAAAA	3652
Db	3571	GCAGTCAAAGGCTTCAAGTTTCTCAGAGCGGAGCGAGTACAAATTGGGACAGTCTAGGA	3630
QY	3653	CTGCGTACTTGGACCCCTGTTATTCGGACAGACGCGTTTCTTAAACAAAAACAGACAA	3712
Db	3631	CTGCGTTCCTGGACCCCTGTATTCGGACAGACGAGTATCAAAAGCATCTGGCGATTAA	3690
QY	3713	CAACAGCAATTTTACTGGACTGGGTCTTCAAAATATTAACCTCAATGGCGCTGATCCAT	3772
Db	3691	CAACAGTGAATCTGTGTGGACTGGAGTCAACAGTCAACCTCAATGGCAAGACTCTCT	3750
QY	3773	CATCAACCCCTGGCACTGCTATGCGCTCAACAAAGACGACGAAAGCAAGTCTTTCCAT	3832
Db	3751	GGTAATC-----CGGCATATGGCAAGCAACAGGACGATGAAGAAAGTTTTTCTCA	3804
QY	3833	GAGCGGTGTCATGATTTTGGAAAAAGAGCGCCGAGACTTCAAAACACTGATTTGACAA	3892
Db	3805	GAGCGGGTCTCATCTTTGGGAACAAAGGCTCAGAGAAAACAATGTGAACATTGA	3864
QY	3893	TGTATGATTTACAGACGAAAGAGAAATTTAAAGCCACTAAACCTGTGGCCACCGAAGATT	3952
Db	3865	GGTCAATGATTTACAGACGAAAGAGAAATCGAAACAACAACTCCGCGCTACCGAGACAT	3924
QY	3953	TGGAACCGTGGCAGTCAATTTCCAGAGACAGACACACCTGTGGACCCGAGATGTGA	4012
Db	3925	TGGTTCTGTATCTACCAACTCCAGAGAGGCAACGAAAGTACCTCGAGATGTCAA	3984
QY	4013	TGCTATGGGACATTAACCTGGCATGATGTGGCAAGATGAGACGTATCTGCAAGGCTCC	4072
Db	3985	CACACAAAGGCTTCTTCCAGGCAATGCTGTGGCAGGACAGAAATGTATCTTCAGGGCC	4044
QY	4073	CATTGGGCCAAATTTCTCAACAGATGACATTTCAACCTGTCTCTTTATGGCGGG	4132
Db	4045	CATCTGGGCAAAAGTTCCACACCGGACGGAATTTTCAACCTCTCCCTTCATAGGGGG	4104
QY	4133	CTTTGGAATCAAGAACCGCGCTCCCAATCCTATCAAAAACAGCGCTGTTCCGCGAA	4192
Db	4105	ATTGGGAATTAACACCTCTCTCCACAAATTTCTATCAAGAACACCCCGGATCTGCGAA	4164
QY	4193	TCTTCGCGGAGTTTCAAGCTAACAAAGTTTGTTCATTATCAACCCAAATCTCCACAG	4252
Db	4165	TCTTCGACACCTTCAGTGGCGGCAAAATTTGCTTCTTCAACACAGTACTCCACGGG	4224
QY	4253	ACA-AGTGAATGTGGAATTTGAATGGGAGCTTGCAGAAAGAAACAGACAGGCTTGGATC	4311
Db	4225	ACACGGTCAAGGTGAGATGCAATGGGAGCTGCAAGAAAGAAACAGAAAGCTTGGATTC	4284
QY	4312	CCGAAGTGCATACATCCCAATTATGCAAAATCTGCAACGTTGATTTTACTGTGACAA	4371
Db	4285	CCGAATTCATGACCTTCCAACTTCAACAAAGTCTGTTAATGTGGAATTACCGTGATTA	4344
QY	4372	ACAATGACCTTATATGAGCCTGGCCCAATTGGGACCCCGTAACCTTACCCGCTCCCTGT	4431

Db	4345	CTAATGGCGGTATATTGAGCGCTGGCCCATTTGSCACAAGATACCTGATCTGTAATCTGT	4404
QY	4432	AATTACGTGTTAATCAATTAACCGGTGATTGCTTTCAAGTTGAACCTTGGTCTCTCTGCC	4491
Db	4405	AATTGCTGTGTAATCAATTAACCGTTTAATTCGTTTCAAGTGAACCTTGGCTCTGCGTA	4464
QY	4492	TTCCTATCTTAATGCGGTACCATGGTTNAGCTTACACATTAACAGCTTGGTTGGCGCTTCG	4551
Db	4465	TTTCTTTCTTATCTAGTTCATAGCTGACGTACGTAGATAGTAGACATGGCGGGTTAATCATTA	4524
QY	4552	CGATTAAGAATTACGTCAATCGGGTTAACCCCTAGTAGTAGAGTTGCCACTCCCTCTCTG	4611
Db	4525	ACTACAAAGGA-----ACCCCTAGTAGTAGAGTTGGCCACTCCCTCTCTG	4568
QY	4612	CGCGCTGCTTCGCTCGGTGGGGCTTGCGGACCAAAAGTCCGCAACGCGCAAGCTTCGT	4671
Db	4569	CGCGCTGCTTCGCTCACTGAGGCCGGCGCGACCAAAAGTTCGCCACCGCCCGGCTTGTGCC	4628
QY	4672	CTGCGCGGCCCAACGAGCGAGCGAGCGCGACAGAGGGAAGTGGGCAA	4718
Db	4629	CGGCGCGCCTTCAGTAGCGAGCGACCGCGCGACAGAGGGAAGTGGGCAA	4675

```

RESULT 8
US-09-782-378A-2
Sequence 2, Application US/09782378A
Patent No. US20020102731A1
GENERAL INFORMATION:
APPLICANT: Hearing, Patrick
APPLICANT: Bahou, Madie
APPLICANT: Sandalon, Ziv
APPLICANT: Gatenko, Dmitri
TITLE OF INVENTION: Adenoviral Vectors
FILE REFERENCE: STONYB-04970
CURRENT APPLICATION NUMBER: US//09/782,378A
CURRENT FILING DATE: 2001-02-12
PRIORITY APPLICATION NUMBER: 60/237,747
PRIORITY FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 4675
TYPE: DNA
ORGANISM: Human adeno-associated virus 2
US-09-782-378A-2

Query Match          64.8% Score 3055.8; DB 11; Length 4675;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY      1 TTGGCCAATCCCTCTCTGCGGCGCTCGCTCGCTCGGTGGGGCTTGCAGACCAAGGTCGC 60
DB      1 TTGGCCATCCCTCTCTGCGGCGCTCGCTCGCTCACTAGAGCCGGGACCAAGGTCGCC 60

QY      61 AGAGCGAGAGCTGTGCTCTGCGGGCCCCACCGAGCGAGCGGCGCGAGAGGAGTGC 120
DB      61 CGAGCGCCGGGCTTTGGCCCGGGCGGCTCTCAGTGACGAGCGAGCGGAGAGGAGTGC 120

QY      121 GCGAAGCTCCATCACTAGAGGGGTAATCGGGAAGCGCTCCACAGCTGCGCGCTCAGCGCTGA 180
DB      121 GCGAAGCTCCATCACTAGAGGGT-----TCTTGAGAGGGGTGAATCTGTA 163

QY      181 CGTAATTTAAGCTAAGGG---GAGTGGTCTCTGTAATTAAGCTGTCAAGTGAAGCTTTTGC 237
DB      164 CGTAATTTAAGCTAAGGGTTAAGGAGAGTCTCTGTAATTAAGGTCAAGTGAAGT---TTTGC 222

QY      238 GACATTTTGGACACACGTTGGCCATTTAGGGTATATATATGCGCGAGTGAAGCAGAGAT 297
DB      223 GACATTTTGGACACACATGTGGTCAAGCTGGGTATTTAAGCCGAGTGAAGCAGCAGAGGT 282

QY      298 CTCATTTTG-AACCGAAATTGAAGAGAGCAGCATGCCGGGCTTTACGAGATCG 356
DB      273 CTCATTTTG-AACCGAAATTGAAGAGAGCAGCATGCCGGGCTTTACGAGATCG 356

```

Db 283 CTCATTTTGAGGCGGAGGTTTGAAGCGCAGCCGCAATGCCGGGTTTTTACAGAGATTG 342
Qy 357 TGAATCAAGTGGCGAGGAGCCTTGAGCAGAGCACTGGCCGGGATTTTTCAGCTCGTTTGA 416
Db 343 TGAATTAAGTCCCGAGGAGCTTGAACGAGCATCTGCCCCGATTTTTCAGACGCTTTTGA 402
Qy 417 GCTGGTGGCGGAGAGAAATGGAGCTGCCCCCGATTTTGAATGATCTGAATCTGA 476
Db 403 ACTGGGTGGCGGAGAGAAATGGAGTTCGCGCAATTTTCAGATGATCTGAATCTGA 462
Qy 477 TTGAGCAGGCAACCTTGAACGTGGCCGAGAACTGACGCGGATCTTGGTCCATGGC 536
Db 463 TTGAGCAGGCAACCTTGAACGTGGCCGAGAACTGACGCGGATCTTGGTCCATGGC 522
Qy 537 GCGCGGAGTAAAGGCGCGGAGGCGCTCTTTTTCAGTTTTCAGTTCGAGAGGCGAGCT 596
Db 523 GCGGTGAGTAAAGGCGCGGAGGCGCTCTTTTTCAGTTCGAGAGGCGAGGAGCT 582
Qy 597 ACTTCCACCTCATATTCGTGTGAGAGCAGCGGGGTCAATCCATGTGCTGGGCGCT 656
Db 583 ACTTCCACATGACGCTGCTGTGAGAAACACCGGGGTGAAATTCATGTGTTTGGAGCTT 642
Qy 657 TCTGAGTCAAGATTAGGAGCAAGCTGTGACAGCATCTACCGGGGATGAGCCGAC 716
Db 643 TCTGAGTCAAGATTAGGAGCAAGCTGTGACAGCATCTACCGGGGATGAGCCGACTT 702
Qy 717 TGCACCATGTTTGGCGGTGACCAAGACGGTATGCGCGCGGAGGGGAGCAAGGTG 776
Db 703 TGCACCATGTTTGGCGGTGACCAAGACGGTATGCGCGCGGAGGGGAGCAAGGTG 762
Qy 777 TGAAGAGTGTACATCTCCCACTACCTCTGAGCCAGACCTGACGCGGATGAGTGG 836
Db 763 TGAAGAGTGTACATCTCCCACTACCTCTGAGCCAGACCTGAGCTCCAGTGG 822
Qy 837 CGTGAATTAATGAGAGAGTATTAAGCGCTGTTTGAACCTGGCCGAGCGCAACGCG 896
Db 823 CGTGAATTAATGAGAGAGTATTAAGCGCTGTTTGAACCTGAACGAGTAAAGGT 882
Qy 897 TCGTGGGCGAGCACTTACCCAGCTGACGAGACCCGAGGAGAGAAACAAGAGAAATGGA 956
Db 883 TGGTGGGCGAGCACTTACCCAGCTGACGAGAGCGAGGAGCAAAAGAGAAATGGA 942
Qy 957 ACCCAATTTCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTACATGAGAGTGG 1016
Db 943 ATCCCAATTTCTGATGCGCGCTGTCTATCAAGTCAAAAACCTTACGCGAGTACATGAGTGG 1002
Qy 1017 TCGGGTGGCTGTGAGACCGGGGATCACTTCGAGAAAGAGTGAATCAGAGAGCAAG 1076
Db 1003 TCGGGTGGCTGTGAGACCGGGGATCACTTCGAGAAAGAGTGAATCAGAGAGCAAG 1062
Qy 1077 CTTCTGATCAATCTCTTCAACGCGCTTCCAACTTGGCGGTCCAGATCAAGCGCGCTTGG 1136
Db 1063 CTTCTGATCAATCTCTTCAACGCGCTTCCAACTTGGCGGTCCAGATCAAGCGCTTGG 1122
Qy 1137 ACAATGCGGAGAGATCATGAGCGCTGACCAATCGCGCGCGACTACCTGATGAGGCCCG 1196
Db 1123 ACAATGCGGAGAGATCATGAGCGCTGACCAATCGCGCGCGACTACCTGATGAGGCCCG 1182
Qy 1197 CTCGCGCGCGGAGCAATTAACCAACCGCATCTACCGCATCTGAGAGTGAACGGCTAGC 1256
Db 1183 AGCCCGGAGAGCAATTTCCAGCAATCGAGATTTTAATAATTTTGAATTAACGGGTAAGC 1242
Qy 1257 AACTGCTTACGCGCGCTCTCTTCTGCGCTGGGCGCAAAAAGTTTGGGAGAGCGCA 1316
Db 1243 ATCCCAATTAAGCGCTTCTCTTCTGAGATGGGCGCAAAAAGTTTGGGAGAGCGCA 1302
Qy 1317 ACAACATCTGAGCTGTTGGCGCGGACACACGCGGAGAGCAACAATGCGGAGAGCGATG 1376
Db 1303 ACAACATCTGAGCTGTTGGCGCGGAGCAATGAGGAGAGCAACAATGCGGAGAGCGATG 1362
Qy 1377 CCGACGCGTGGCTTCTTACGCGTGGGTCAAATGAGCAATTAAGCAATTTCTCTCATG 1436
Db 1363 CCGACATGTCCTTCTTACGCGGTGGTAAATGAGCAACAATTAAGCAATTTCTCTCATG 1422

Qy 1437 ATTGCGTCAACAAGATGTGATCTGGTGGGAGGAGGAGCAAGATTAAGCGGCAAGGTG 1496
Db 1423 ACTGTGTCAACAAGATGTGATCTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1482
Qy 1497 AGTCCGCAAGGCAATTTCTGGCGGAGCAAGAGTGGCGGTGAGCAAAAGTGAAGTCT 1556
Db 1483 AGTCCGCAAGGCAATTTCTGGAGGAGCAAGAGTGGCGGTGAGCAAAAGTGAAGTCT 1542
Qy 1557 CCGCCAGATGACCCCAACCCCGTGAATGTCTCACTTCAACCAACCAATGTGGCGGTGA 1616
Db 1543 CCGCCAGATGACCCCAACCCCGTGAATGTCTCACTTCAACCAACCAATGTGGCGGTGA 1602
Qy 1617 TTGACGGGAAACACACCACTTGAACACAGAGCGGTTGAGAGCCGGAATGTAAT 1676
Db 1603 TTGACGGGAAACACACCACTTGAACACAGAGCGGTTGAGAGCCGGAATGTAAT 1662
Qy 1677 TTGAATCAACCCGCTGTGAGACATGACTTTTGGCAAGTGAACAAGAGAGAGTCAAG 1736
Db 1663 TTGAATCAACCCGCTGTGAGACATGACTTTTGGAAAGTCAACAAGAGAGAGTCAAG 1722
Qy 1737 AGTTCTTCCGCTGGCGAGAGATCACTGAACGAGGTGGCGATGAGTTCTACGTAGAA 1796
Db 1723 AGTTCTTCCGCTGGCGAGAGATCACTGAACGAGGTGGCGATGAGTTCTACGTAGAA 1782
Qy 1797 AGGTTGAGGCAACAAGAGCCGCGCCGAGTGAACGCGATTAAGCGAGCCCAAGCGG 1856
Db 1783 AGGTTGAGGCAACAAGAGCCGCGCCGAGTGAACGCGATTAAGTGAAGCCCAAGCGG 1842
Qy 1857 CCGCCCTTCACTGCGGAGATCCATGACGTCAAGCGGAGAGAGGCTCGGTGACTTTG 1916
Db 1843 TGGCGAGTCAAGTTGGGAGGAGCATGACGTCAAGCGGAGAGGCTCGGTGACTTTG 1899
Qy 1917 CCGAGAGTCAACAAGAGTGTCTGTCAGCGGCGATGCTTCAAGTCTGTTCCCT 1976
Db 1900 CAGACAGTCAACAAGAGTGTCTGTCAGCGGCGATGAGTCTGTTCCCT 1959
Qy 1977 GCAAGACATGCGGAGAGAGAGTGAATCAAGATTTTCAACTTGTCTTCAAGCGAGAG 2036
Db 1960 GCAAGACATGCGGAGAGAGAGTGAATCAAGATTTTCAACTTGTCTTCAAGCGAGAG 2019
Qy 2037 ACTGTTCAAGTGTCTTCCCGCGGTGAGAACTTCAACCGGT---GTCAGAGAGAG 2093
Db 2020 ACTGTTCAAGTGTCTTCC---GTCAGAGATCTCAACCGGTGTCGTCAAGAG 2076
Qy 2094 CGTATCGAGAACTGTGTCCATTCATCTGTGTGGGGCGGGCTCCGAGATGCTTGT 2153
Db 2077 CGTATCGAGAACTGTGTCCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2133
Qy 2154 CGGCTTCCGATCTGTGTCACAGTGGAGCTGTGATCTGTGTTTGAAGCAATTAAGT 2213
Db 2134 CGGCTTCCGATCTGTGTCACAGTGGAGTGTGATCTGTGATCTGTGTTTGAAGCAATTAAGT 2193
Qy 2214 AAACAGATATGAGCTCCGATGAGTATCTTCCAGATTTTGGTGAAGAGCAACTCTGAG 2273
Db 2194 AAATCAGATATGAGCTCCGATGAGTATCTTCCAGATTTTGGTGAAGAGCAACTCTGAG 2253
Qy 2274 GGCATTCGAGTGTGAGGACTTGAACCTGAGCCCGAGAGCCCAAGCCAGAGCA 2333
Db 2254 GGAATTAAGAGAGTGTGAGAGCTCAAACTGGCCACACACCAAGCCCGAGAGCGG 2313
Qy 2334 AAGCAGAGAGAGCGCGGGGTCTGTGTTCTTCTGAGTCAAGATCTTGGAGCCCTTCAAC 2393
Db 2314 CATTAAGAGAGAGAGGGGTCTTGTGTTCTTCTGAGTCAAGATCTTGGAGCCCTTCAAC 2373
Qy 2394 GGAATCAAGAGAGAGCCCGTCAACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2452
Db 2374 GGAATCAAG 2433
Qy 2453 GGCCTACAGACAG 2512
Db 2434 AGCTTACAG 2493

QY	2513	GGCCGACGTTTACAGAGCGCTCTGCAAGAAAGATACGTTTGGGGGCAACCTCGGGCAGC	2572
Db	2494	CGCCGAGTTTACGAGCGCCTTTAAAGAAAGATACGTTTGGGGGCAACCTCGGACGAGC	2553
QY	2573	AGTCTTCCAGGCGCAAGAGCGGGTCTCGAACTCTCGGTCTGGTTGAGGAAGCGCTTAA	2632
Db	2554	AGTCTTCCAGGCGAAAAAGAGGGTCTTGAACTCTCGGGCTGGTTAGGAACCTGTTTAA	2613
QY	2633	GACGAGCTCTGGAAAGAAACGTCGCGGTAGACAGTCGACAGAAGCCAGACTCTCTCTC	2692
Db	2614	GACGAGCTCTGGGAAAAAAGAGGCGGGTAGAGACACTCTCTGTGGAGCCAGACTCTCTCTC	2673
QY	2693	GGGCAATCGGCAAGACAGCCGACAGACCCCGCTAAAAAGAGACTCAATTTTGGTCAACCTG	2752
Db	2674	GGGAAACCGGAAAGGCGGGCCAGACAGCCCTGCAAGAAAAAGATTGAATTTTGGTCAAGACTGG	2733
QY	2753	CGACTCAGAGTCAGTCCCCGATCCACAACCTCTGGAGAACTTCAGCAACCCCGCGTCG	2812
Db	2734	AGAGCGAGACTCAGTACTGACCTGACCCCCAGCCTCTGCGACAGCCACCGAGCCCTCTCGG	2793
QY	2813	TGTGGGACCTTACTACATGGCTTCAGGCGGTGGCGCACCAATGGCAGACCAATTAACGAAG	2872
Db	2794	TCTGGGAACCTAATACATGGCTCTACAGGCGAGTGGCGCACCAATGGCAGACCAATTAACGAAG	2853
QY	2873	CGCCGACGGAGTGGGTAAATGCTCTCAGAGAAATTTGGATTTGCGATTTCCACATGGCTGGGGA	2932
Db	2854	CGCCGACGGAGTGGGTAAATTTCTCGGAAATTTGGATTTGCGATTTCCACATGGATGGGGA	2913
QY	2933	CAGAGTCATCAACACACGACCCGCACTGGGCGCTTGGCCACTTGCACATTAACCACTCTTA	2992
Db	2914	CAGAGTCATCAACACACGACCCGCACTGGGCGCTTGGCCACTTGCACATTAACCACTCTTA	2973
QY	2993	CAAGCAATCTCCAGTGGCTTCAACGGGGGCGACGAACGACACCATTACTTGGCTACAG	3052
Db	2974	CAACCAATTTTCCAGCAATCA--GAGACCTCGAAGCAACATCACTACTTTGGCTACAG	3030
QY	3053	CACCCCTCGGGGGATTTTGAATTTAAACAGATTCACATGCCACTTTTACAACAGGATCG	3112
Db	3031	CACCCCTTGGGGGATTTTGAATTTCAACAGATTCACATGCCACTTTTACAACAGGATCG	3090
QY	3113	GCAGGACTCATCATCAACAATTTGGGGATTCGGGCGCCAAAGAGACTCAACTTCAAACTCTT	3172
Db	3091	GCMAAGACTCATCATCAACAATTTGGGGATTTCCGACCCAAAGAGACTCAACTTCAAGCTCTT	3150
QY	3173	CAACATCCAGTCAAGAGAGTCAACGACGAAATGATGGCGTCAACCAATCGCTAATTAACCT	3232
Db	3151	TAACTTCAAGTCAAAAGAGGTACGACAGATGACGATGACAGATTTGCCAATTAACCT	3210
QY	3233	TACCAACACGTTTCAAGTCTTCTCGGACTCGGAGTACCAAGTTCCTGATCGTCTCGGCTC	3292
Db	3211	TACCAACACGTTTCAAGTGTGTTTACTGACTCGGAGTACCAAGTTCCTGATCGTCTCGGCTC	3270
QY	3293	TGGCGACACAGGGCGCTCCCTCGGTCGCTTCCGGCGGACGCTTCAATGATTCGGCAATACG	3352
Db	3271	GGCGCATAAAGATCTCTCCGCGGTTTCCACAGACGCTCTTCAATGATGACCAAGTATG	3330
QY	3353	CTACCTGACGCTCAACAATGGCAGCAAGCCGTGGAGCTTCACTCTTTTACTGCTGGA	3412
Db	3331	ATACTCAACCTGAAACAAAGGAGTCAAGGAGTACAGAGCTTCAATTTTACTGCTGGA	3390
QY	3413	ATATTTCCCTTCTCAGATGCTGAGAAAGGGCAACAATTTAACCTTCACTACACTTTTGA	3472
Db	3391	GTACTTTCTCTTCTCAGATGCTGAGTACCGGAAACAATTTAACCTTCACTACACTTTTGA	3450
QY	3473	GGAGTGGCTTTCCACAGAGTACAGGCAACGACGAGGCCGTGACCGGGCTGATGAATCC	3532
Db	3451	GGAGCTTCTCTTCCACAGAGTACAGGCTTCAAGCGAGTCTGACCGCTTCAATTAATCC	3510
QY	3533	TCTCATGACCAATACCTGTATTAACCTGAACGAATCAAAATCAAGTCCGGAAGTCCCA	3592
Db	3511	TCTCATGACCAAGTACTGTATTAACCTTACGAGAAACAACAATCAAGTGGAAACAACAC	3570
QY	3593	AAACAGAGACTTGGCTGTTTACCGGTGGCTTCAAGTGGCAATGTCGTTTCAACCCAAAAA	3652

Db	3571	GCAGTCMAAGGCTTAGTTCCTCAGGCCGAGCCAGTGAACATTGGGGAACAGTCTTAGGAA	3630
Qy	3653	CTGGCTACCTGGACCTCTGTTATCGGCAGACGCGGCTTTCTAATAACAAACACAGCAACA	3712
Db	3631	CTGGCTTCCTGGACCTCTGTTATCCGCCAGACGCGAGTATCAAAAGACATCTGGCGATTAACA	3690
Qy	3713	CAACAGCAATTTTACCTGGACCTGGTCTCTCAAAATATTAACCTCAATGGCGGTGAATCCAT	3772
Db	3691	CAACAGTGAATTAACCTGGACCTGGAGCTTACCAATATCCACTCAATGGCAAGACTCTCT	3750
Qy	3773	CATCAACCCCTGGACCTGATGGCTCACAACAAGACGACGACAAACAAAGTTCTTCCCAT	3832
Db	3751	GGTGAATC-----CGGCCATGGCAAGCCACACAGACGATGAAGAAAAAGTTTTTCCCTCA	3804
Qy	3833	GAGCGGTGCATGATTTTTTGGAAAAAGAGCGCGGAGCTTCAAAACACTGACATTGGACAA	3892
Db	3805	GAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAATATGAAACATTGAAAA	3864
Qy	3893	TGTCTATGATTTACAGACGAAAGAGAAATTTAAAGCACTTAACCTTTGGCCACCCGAAAGTT	3952
Db	3865	GGTATATGATTTACAGACGAAAGAGAAATTCGAAACAACATCTCCGTGTACGGAGCAGTA	3924
Qy	3953	TGGACCGCTGGCAGTCAATTTCCAGAGCAGACAGACCCCTCGACCCGAGATGTGCA	4012
Db	3925	TGGTTCTGTATCTACCAACTTCAGAGAGGCAACAGACAGACTACCGAGATGTCA	3984
Qy	4013	TGCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACCTGCAGGCTCC	4072
Db	3985	CACACAAGGCGTTCTTCAGGCGATGTGTGGCAGAGACAGATGTGTACTTTCAGGGGCG	4044
Qy	4073	CATTTGGGCCAAATTCCTCAACAGATGACACTTTTCAACCCGCTCTCTTATGGGCG	4132
Db	4045	CATCTGGCCAAAGATTCACACAGAGCGGACATTTTCAACCCCTCTCCCTCATAGGGTGG	4104
Qy	4133	CTTTGGAATCAAGAACCGGCTCTCTCAGATTCCTCATCAAAAAACAGCGCTGTCTCGCAA	4192
Db	4105	ATTGGACTTAAACACCTCTCTTCCACAGATTTCTATCAAGAACACCCCGTACTCGCAA	4164
Qy	4193	TCTCTCCGCGAGTTTCACTACGTAACAAAGTTTGCTTCAATTCATCAACCAATATCTCCACGG	4252
Db	4165	TCTTTGACCACTTCAGGTGGGAAAGTTTGCTTCTTCAATCAACAGTACTCCAAGGG	4224
Qy	4253	ACA-AGTAGTGTGAAATTTGAATGGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGGAATC	4311
Db	4225	ACACGGTCAAGGTGAGATTCAGTGGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGGAATC	4284
Qy	4312	CGAAGTGCAGTACACATCCAAATATGACAAATCGCCAAAGTTGATTTTACTGAGACA	4371
Db	4285	CGAAATTCAGTACCTTCCAACTACACAAAGTCTGTTATATCGTGAATTAACCTGAGATA	4344
Qy	4372	ACAATGCACTTATACAGGCTCGCCCATATGGCAACCCGTTAATCCCGTCCCTGT	4431
Db	4345	CTAATGGGCTATATCAGAGCGCTCGCCCATATGGCAACAGATTAATCTGTAATCTGT	4404
Qy	4432	AATTAAGTGTAAATCAATAAACCGGTTGATTTGCTTCACTTGAACCTTGGTCTCTGTCC	4491
Db	4405	AATGCTGTAAATCAATAAACCGTTTAAATCGTTTCAGTTGAACCTTGGTCTCTGCGTA	4464
Qy	4492	TTCTATCTTAATCGGTTACAGATGTATAGCTTACACTTAACGCTTGGTGGGCTCG	4551
Db	4465	TTCTCTTTATTAATGATTTCAATGCTACGTAGATTAAGTACATGGCGGGTTAATCATTA	4524
Qy	4552	CGATTAAGACTTACGTCATCGGGTTAACCCCTAGATGAGATTGCCACTCCCTCTCG	4611
Db	4525	ACTACAAAGA-----ACCCCTAGATGAGATTGGCACATCCCTCTCTG	4588
Qy	4612	CGCGCTGCTGCTCGGTGGGGCTGGGGAACAAAGTTCGCGACGCGCAGAGCTTGCT	4671
Db	4569	CGCGCTGCTGCTCACTGAGGCGGGGACAAAGGTGCGCCGACGCGCGGCTTGGCC	4628
Qy	4672	CTGCCGCGCCACCGAGGACCGAGCGCGCAGAGGGAGTGGGCAA	4718

Db 1723 ACTTTTCGGTGGCAAGGATCACTGTTGAGGTGAGCATGATTTCTACGTCAAA 1782
Qy 1797 AGGTGGAGCCAAAGAACCCGCCCGCATGACGGGATTAAGAGAGCCCAAGGGG 1856
Db 1783 AGGGTGGAGCCAAAGAACCCGCCCGCATGAGAGATTAAGAGAGCCCAAGGGG 1842
Qy 1857 CCGGCTTCAGTGGCGGATCCATGACGATGAGAGCGGAGAGAGCTCCGGTGAATTG 1916
Db 1843 TGGCGAGTCACTTGGCCGAGCATGACGATGAGAGCGGAGAGAGCTCCGGTGAATTG 1899
Qy 1917 CCGAGAGTACCAAAACAAATGTTCTGCTCAAGCGGCGATCTTCAAGTGTGTTCCCT 1976
Db 1900 CAGACAGGTACAAACAAATGTTCTGCTCAAGTGTGTTCCCT 1959
Qy 1977 GCAGACATGCGAGAGATGATGATCAATTTCAATTTGCTTCAAGAGAGAGAG 2036
Db 1960 GCAGACATGCGAGAGATGATGATCAATTTCAATTTGCTTCAAGAGAGAGAG 2019
Qy 2037 ACTGTTGAGAGTGTCCCGGCGGTGAGATCTCAAGCGGT---CGTCAAGAGAGAG 2093
Db 2020 ACTGTTGAGAGTGTCCCGGCGGTGAGATCTCAAGCGGT---CGTCAAGAGAGAG 2076
Qy 2094 CGTATCGGAACTCTGTGCTCATCTGCTGAGGCGGCTCCGAGATGCTTCT 2153
Db 2077 CGTATCGGAACTCTGTGCTCATCTGCTGAGGCGGCTCCGAGATGCTTCT 2133
Qy 2154 CGGCTGCGATCTGTGCTCATCTGCTGAGGCGGCTCCGAGATGCTTCT 2213
Db 2134 CTGCTGCGATCTGTGCTCATCTGCTGAGGCGGCTCCGAGATGCTTCT 2193
Qy 2214 AAACAGGATGCGTGGCGGATGATGATCTCAAGTGTGCTGAGAGAGAGAG 2273
Db 2194 AAATCAAGTATGCTGCTGCTGATGATGATCTCAAGTGTGCTGAGAGAGAGAG 2253
Qy 2274 GGCATTGCGAGTGTGAGATTTGAACTGAGAGAGAGAGAGAGAGAGAGAGAG 2333
Db 2254 GGAATTAAGACAGTGTGAGAGATTTGAACTGAGAGAGAGAGAGAGAGAGAGAG 2313
Qy 2334 AAGCAG 2393
Db 2314 CATTAAG 2373
Qy 2394 GGCATTGCGAGTGTGAGATTTGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2452
Db 2374 GGAATTAAGACAGTGTGAGAGATTTGAACTGAGAGAGAGAGAGAGAGAGAGAGAG 2433
Qy 2453 GGCATTGCGAGTGTGAGATTTGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2512
Db 2434 AGCCTAAG 2493
Qy 2513 CGCGAGATTTGAG 2572
Db 2494 CGCGAGATTTGAG 2553
Qy 2573 AGTCTTCAAG 2632
Db 2554 AGTCTTCAAG 2613
Qy 2633 GAGCGCTCTGAG 2692
Db 2614 GAGCGCTCTGAG 2673
Qy 2693 GGGCATGCGAG 2752
Db 2674 GGGCATGCGAG 2733
Qy 2753 CGACATGAG 2812
Db 2734 AG 2793
Qy 2813 TGTGGAG 2872

Db 2794 TGTGGAG 2853
Qy 2873 CGCGAG 2932
Db 2854 CGCGAG 2913
Qy 2933 CAGAGATCATCAG 2992
Db 2914 CAGAGATCATCAG 2973
Qy 2993 CAGAGATCATCAG 3052
Db 2974 CAGAGATCATCAG 3030
Qy 3053 CAGAGATCATCAG 3112
Db 3031 CAGAGATCATCAG 3090
Qy 3113 CAGAGATCATCAG 3172
Db 3091 CAGAGATCATCAG 3150
Qy 3173 CAGAGATCATCAG 3232
Db 3151 CAGAGATCATCAG 3210
Qy 3233 CAGAGATCATCAG 3292
Db 3211 CAGAGATCATCAG 3270
Qy 3293 CAGAGATCATCAG 3352
Db 3271 CAGAGATCATCAG 3330
Qy 3353 CAGAGATCATCAG 3412
Db 3331 CAGAGATCATCAG 3390
Qy 3413 CAGAGATCATCAG 3472
Db 3391 CAGAGATCATCAG 3450
Qy 3473 CAGAGATCATCAG 3532
Db 3451 CAGAGATCATCAG 3510
Qy 3533 CAGAGATCATCAG 3592
Db 3511 CAGAGATCATCAG 3570
Qy 3593 CAGAGATCATCAG 3652
Db 3571 CAGAGATCATCAG 3630
Qy 3653 CAGAGATCATCAG 3712
Db 3631 CAGAGATCATCAG 3690
Qy 3713 CAGAGATCATCAG 3772
Db 3691 CAGAGATCATCAG 3750
Qy 3773 CAGAGATCATCAG 3832
Db 3751 CAGAGATCATCAG 3804
Qy 3833 CAGAGATCATCAG 3892
Db 3805 CAGAGATCATCAG 3864
Qy 3893 CAGAGATCATCAG 3952
Db 3865 CAGAGATCATCAG 3924

QY 3953 TGGAGCGGTGAGCAATTTCCAGAGCAGACAGACCCCTGCGACCGGAGATGTCA 4012
 DB 3925 TGGTTCGTATCTACCACTCCAGAGAGGCAACAGACAGACCTACCCGAGATGTCA 3984
 QY 4013 TGTATGAGAGCAATTAACCTGAGCATGTGTGCAAGATAGACCTGTACTCTGAGGGTCC 4072
 DB 3985 CACACAGAGCGTCTTCAGGAGATGTGTGCAAGAGATGTGTACTCTGAGGGGCC 4044
 QY 4073 CATTTGGGCAAAATTTCTCAGACAGATGACACTTTCAACCCGCTCTCTTATGGGCG 4132
 DB 4045 CATCTGGGCAAAAGTTCCACACAGAGGACATTTTCACCCCTCTCCCTCATGGGTGG 4104
 QY 4133 CTTTGAGCTCAAGAACCGCTCTCTCAGATCTCATCAAAAACAGCGCTGTTCTGAGAA 4192
 DB 4105 ATTGGAGCTTAACACCTCTCTCAGACATTTCTCATAGAACCCCGGATCTGAGAA 4164
 QY 4193 TCTCCGCGGAGTTTCACTCAAAAGTTTGTCTTCAATTCATCAACCAATCTCAGAG 4252
 DB 4165 TCTTTCAGCACCTTCACTGAGTGGGCAAGTTTGTCTTCAATCAACAGATCTCAGGG 4224
 QY 4253 ACA-AGTGTGTGAAATTTGAATGAGAGCTGCGAGAAAGAAACAGCAAGCGTGGATC 4311
 DB 4225 ACAGGTCAGGTCAGATGAGTGGAGCTGCGAGAAAGAAACAGCAAGCGTGGATC 4284
 QY 4312 CCGAAGTGAAGTACATCCAAATTAATGCAAAATCTGCAAGCTTGAATTTTATCTGAGCA 4371
 DB 4285 CCGAATTCAGTACATCTTCACTACACAAAGTCTGTTAATCTGAGACTTACCGTGAATA 4344
 QY 4372 ACAATGACTTATATAGAGCTGCGCCCATTTGACACCCGTTACCTTACCCGCTGCT 4431
 DB 4345 CTATGAGGAGTATTCAGAGCTGCGCCCATTTGACACCAAGATACCTGATCTGATGTGT 4404
 QY 4432 AATTAGCTTTAATCAATTAACCGGTTGATTTGTTTCAAGTTGAACCTTGTCTCTGTC 4491
 DB 4405 AATTGCTTTAATCAATTAACCGTTAATGTTTCAAGTTGAACCTTGTCTCTGTC 4464
 QY 4492 TTCTATCTTATGAGTACCAATGTTAAGCTTACATTAACCTTGTGTTGCTGCTG 4551
 DB 4465 TTCTTTCTTATCTGATTTTCATGCTACAGTATGATGATGATGAGCGGTTAATCAATA 4524
 QY 4552 CGATTAAGACTTACGTCATGAGGTTACCCCTAGTGTGAGTGTGCCCATCTCTCTG 4611
 DB 4525 ACTAAGAGG-----ACCTTATGATGAGTGTGCCCATCTCTCTG 4568
 QY 4612 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4671
 DB 4569 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4628
 QY 4672 CTGCGGCGCCCAAGAGCGAGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4718
 DB 4629 CGGCGGCTTCAAGTGAAGAGGAGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 4675

RESULT 10

US-10-205-942-5
 ; Sequence 5, Application US/10205942
 ; Publication No. US2003005390A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of No. US2003005390A1Ch Carolina-Chapel Hill
 ; APPLICANT: Rabinowitz, Joseph E.
 ; APPLICANT: Samulek, Richard J.
 ; APPLICANT: Xiao, Weidong
 ; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
 ; FILE REFERENCE: 5470-186
 ; CURRENT APPLICATION NUMBER: US/10/205,942
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 8179
 ; TYPE: DNA
 ; ORGANISM: Adeno-associated virus

US-10-205-942-5

Query Match 62.4%; Score 2941.8; DB 15; Length 8179;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 3517; Conservative 0; Mismatches 792; Indels 20; Gaps 8;

QY 205 GTCCTGATTTAGCTGTCACGTCGATGTCCTTTTGGACACTTTTGGACACCAAGTGGCCATT 264
 DB 121 GTCCTGATTTAGAGTCACGTCGATGTCCTTTTGGACACTTTTGGACACCAAGTGGTCACG 179
 QY 265 TAGGTAATATATGCGCGAGTGAAGCGAGCAGAGATCTCCATTTTG-ACCGCAAAATTTGAAC 323
 DB 180 CTGGATATTTAAGCCCGAGTGAAGCAGACAGAGGCTCTCATTTTGAAGCGGAGGTTTGAAC 239
 QY 324 GAGCAGAGCAGTACCGGCTTTTGAAGATCTGATCAAGGTGCGGAGGAGCTTGACG 383
 DB 240 GCGAGCGCGCATGCGGCTTTTGAAGATCTGATCAAGGTGCGGAGGAGCTTGACG 299
 QY 384 AGCAGCTGCGGCGATTTTGAATCTGCTTTTGTAGCTGCGGCTGCGAGAGAAATGGAGC 443
 DB 300 GGCATCTGCGGCGATTTTGAATCTGCTTTTGTAGCTGCGGCTGCGAGAGAAATGGAGT 359
 QY 444 TGCCTCCGATTTTGAATGATGATCTGATTTGAAGAGGACCCCTGACCGTGGCG 503
 DB 360 TGCCTCCGATTTTGAATGATGATCTGATTTGAAGAGGACCCCTGACCGTGGCG 419
 QY 504 AGAAGCTGACGCGGCTTCTGCTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 563
 DB 420 AGAAGCTGACGCGGCTTCTGCTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 479
 QY 564 TCTTCTTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 623
 DB 480 TTTTCTTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 539
 QY 624 CCAAGGAGGTCATATCATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 683
 DB 540 CCAAGGAGGTCATATCATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 599
 QY 684 TGCAGACATCTACCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 743
 DB 600 TGCAGACATCTACCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 719
 QY 744 CCGGATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 803
 DB 660 CCGGATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 719
 QY 804 TCTGTCCTAAGATCTACCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 863
 DB 720 TCTGTCCTAAGATCTACCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 779
 QY 864 GCGCTGTTTGAACCTGCGGAGGCAACGCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 923
 DB 780 GCGCTGTTTGAACCTGCGGAGGCAACGCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 839
 QY 924 GCGAGCCGAGAGCAAGCAAGAGATCTGAACCCCAATTTGTCGCGGCTGTATCC 983
 DB 840 GCGAGCCGAGAGCAAGCAAGAGATCTGAACCCCAATTTGTCGCGGCTGTATCC 899
 QY 984 GGTCAAAATCTCTCGGCGCTACATGAGTGTGTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 1043
 DB 900 GATCAAAATCTCTCGGCGCTACATGAGTGTGTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 959
 QY 1044 CTTCCGAGAGCAGTGTATCAGAGAGCAGGCTCTGTCATCTCTTCAAGCGGCTT 1103
 DB 960 CTTCCGAGAGCAGTGTATCAGAGAGCAGGCTCTGTCATCTCTTCAAGCGGCTT 1019
 QY 1104 CCAACTGCGGTCCTCAGATCAAGCGGCTCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1163
 DB 1020 CCAACTGCGGTCCTCAGATCAAGCGGCTCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1079
 QY 1164 CCAATTCGCGGTCCTCAGATCAAGCGGCTCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1223
 DB 1080 CCAATTCGCGGTCCTCAGATCAAGCGGCTCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1139

QY 1224 GCATCTACCGCATCTCTGAGTGAACGGCTACGAACTGCTTACCGCGCTCGCTTTC 1283
Db 1140 GGATTTTATTAATTTTGGAACTTAAACGGGTACGATCCCAATTAAGCGCTTCCGCTTTC 1199
QY 1284 TCGGCTGGGCGCCAGAAAAAGTTTGGGGAAGCGCAACACCATCTGGCTGTTTGGGCGGCA 1343
Db 1200 TGGGATGGGCGACGAAAAAGTTTGGGCAAGGAACACATCTGGCTGTTTGGGCGTGA 1259
QY 1344 CCACGGGCAAGACCAACATCGCGGAGAGCATCGCCCAACGCGCTGCTTCTACGGCTGCG 1403
Db 1260 CTACCGGGAGAACCAACATCGCGGAGGCGCATAGCCCACTATGCTCTTCTACCGGTGCG 1319
QY 1404 TCACTGGACCAATGAGAACTTTCCTTCAATGATGCGTGAACAGATGGGATCTGGT 1463
Db 1320 TAAATGGACCAATGAGAACTTTCCTTCAACGATGCTGTCGAAAGATGGATCTGGT 1379
QY 1464 GGGAGAGGGGCAAGATGACGGCCAAAGTCTGAGAGTCCGCCAAGGCCATTTCTGGCGGCA 1523
Db 1380 GGGAGAGGGGCAAGATGACGGCCAAAGTCTGAGAGTCCGCCAAGGCCATTTCTGGAGGA 1439
QY 1524 GCAGAGTGCCTGAGGCAAAAAGTGCAGAGTCTCGCCCAAGTGAATGACCCCAACCCCTGA 1583
Db 1440 GCAAGGTGCGGTGAGACCAAGAAATGCAAGTCTCGGCCAGATGACCCCACTCCGTGA 1499
QY 1584 TCGTCACTCCAGACCAACATGTCGCGCGGATTTGAACGGGAAACAGACCACTTTCAGC 1643
Db 1500 TCGTCACTCCAGACCAACATGTCGCGCGGATTTGAACGGGAACTCAACGACCTTTCAGAC 1559
QY 1644 ACCAGACGCGCTTGCAGAGACCGGATGTTCAATTTGAATCTACCCGCGCTGAGAGATG 1703
Db 1560 ACCAGACGCGCTTGCAGAGACCGGATGTTCAATTTGAATCTACCCGCGCTGAGATATG 1619
QY 1704 ACTTTGGCAAGTGAACAAAGCAAGAGTCAAAAGATTTCTGCTGGGCGAGATCAAG 1763
Db 1620 ACTTTGGCAAGTGAACAAAGCAAGAGTCAAAAGATTTCTGCTGGGCGAGATCAAG 1679
QY 1764 TGACCGAGGTGGCGATGATTTTCAACGTCAAGAAAGGTGAGACCAACAAAGACCCGCGC 1823
Db 1680 TGCTTGAAGTGAAGATTAATTTCTACGTCAAAAGGTGAGACCAACAAAGACCCGCGC 1739
QY 1824 CCGATGACCGGATTAAGACGAGCCCAAGCGGCGCTGCGCTCACTGCGGATCATGCA 1883
Db 1740 CCGATGACCGGATTAAGATGAGCCCAAGCGGCGCTGCGGATCATGCGAGCATGCA 1799
QY 1884 CGTCAAGACGGGAAGAGACTCCGCTGAGCTTTGCGCAAGGTACCAAAACAAATGTTCTC 1943
Db 1800 CGTCAAGACGGGA--AGCTTCAATCACTACGCAAGAGTACCAAAACAAATGTTCTC 1856
QY 1944 GTCAAGCGGCGATGCTGAGATGCTTTCCCTGCAAGACATGCGAGATGATGATGCA 2003
Db 1857 GTCAAGCGGCGATGATGATGCTTTCCCTGCAAGATGCGAGATGATGATGCA 1916
QY 2004 ATTTCAACATTTGCTTCAACGAGCGAGAGACTGTTCAAGAGTCTTCCCGGCGTGT 2063
Db 1917 ATTTCAACATTTGCTTCACTACGAGCAGAAAGACTGTTTAAAGGCTTTC--CGGT 1973
QY 2064 CAGATCTCAACCGGT--CGTCAAGAAAGAGAGCTATCGAAACTCTGTGCTTCAATC 2120
Db 1974 CAGATCTCAACCGGTCTGTCTGCAAAAGGCTATCAAACTGTGCTCAATCAATC 2033
QY 2121 ATTCGTGGGCGGAGCTCCGAGATGTTGCTTGGCTGCGCTGAGATCTGATCAACGTGAC 2180
Db 2034 ATATCATAGGAAAGGTGCGA---CGTTTCACTGCTGCGATCTGATCAATGATGAT 2090
QY 2181 TGAATGATGATGTTCTGAGCAATTAATGACTTAAACAGGTATGCTGCGATGATGAT 2240
Db 2091 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2150
QY 2241 CTTTCAAGTGGCTCGAGAGCAACCTCTGAGGAGCATTCGAGAGTGGTGGAGCTTGAA 2300
Db 2151 CTTTCAAGTGGCTCGAGAGCACTCTCTGAGAGATTAAGACAGTGGTGGAGCTTAA 2210

QY 2201 CTTGAGACCCCGAAGCCCAAGCCCAACAGCAAAAGAGAGCAGCGCGCGGCTGCTG 2360
Db 2211 CTTGAGACCCCGAAGCCCGAAGCCCGAAGCCCGAAGCCCGAAGCCCGGCTGCTG 2270
QY 2361 CTTTCTGGCTTACAGTACCTTCCGACCTTTCAGAGTCTGCAAGAGGGGAGCCGCTCAAC 2420
Db 2271 CTTTCTGGCTTACAGTACCTTCCGACCTTTCAGAGTCTGCAAGAGGGGAGCCGCTCAAC 2330
QY 2421 GCGGCGGAGCGAGCGGCTGAGGCAAGAGGCTGAGAGAGGCTGAGAGAGGCTG 2480
Db 2331 GAGGAGAGCGCGGCTGAGGCAAGAGGCTGAGAGAGGCTGAGAGAGGCTGAGAGG 2390
QY 2481 GACATCTGATCTGAGGATTAACCAACGCGGAGGCTGAGAGAGGCTGAGAGG 2540
Db 2391 GACATCTGATCTGAGGATTAACCAACGCGGAGGCTGAGAGAGGCTGAGAGG 2450
QY 2541 GATACGCTTTTGGGGCAACCTTGGGCGAGAGTCTTCCAGGCTCAAGAGCGGCTTC 2600
Db 2451 GATACGCTTTTGGGGCAACCTTGGGCGAGAGTCTTCCAGGCTCAAGAGCGGCTTC 2510
QY 2601 GAACCTCTGAGTCTGAGTGAAGAGGCTGAGAGGCTCTGAGAGAGAGGCTGAG 2660
Db 2511 GAACCTCTGAGTCTGAGTGAAGAGGCTGAGAGGCTCTGAGAGAGAGGCTGAG 2570
QY 2661 GAGAGTGCAGCAAGAGCCAGACTCTCTCGGAGATCGGCAAGACAGCCAGACGCC 2720
Db 2571 GAGAGTGCAGCAAGAGCCAGACTCTCTCGGAGATCGGCAAGAGCCAGACGCC 2630
QY 2721 GCTTAAAGAGACTCAATTTTGGTCAAGCTGAGAGCTCAAGTCACTCCGATCCACA 2780
Db 2631 GCAAGAAAGATTAATTTTGGTCAAGCTGAGAGCTCAAGTCACTCCGATCCACA 2690
QY 2781 CTTCTGGAGAACCTTCCAGCAACCCCGCTGCTGAGGACCTTCAATAGGCTTCAAG 2840
Db 2691 CTTCTGGAGAACCTTCCAGCAACCCCGCTGCTGAGGACCTTCAATAGGCTTCAAG 2750
QY 2841 GGTGCGCAACCAATGAGCAATTAAGAGGCGCGAGAGTGGATATGCTTCAG 2900
Db 2751 AGTGGCGCAACCAATGAGCAATTAAGAGGCGCGAGAGTGGATATGCTTCAG 2810
QY 2901 AATTGGCATTTGCAATTTTCAATGAGTGGGCGAGAGTCACTCAACAGACCCGCAAC 2960
Db 2811 AATTGGCATTTGCAATTTTCAATGAGTGGGCGAGAGTCACTCAACAGACCCGCAAC 2870
QY 2961 TGGGCTTGGCCACTTCAATTAACCACTTCAAGAGCAATCTCAAGTCTTCAACGG 3020
Db 2871 TGGGCTTGGCCACTTCAATTAACCACTTCAAGAGCAATCTTCAAGTCTTCAACGG 2927
QY 3021 GCCAGCAACCAACCACTTCTGAGTACAGACCCCTGAGGAGTATTTGATTTCAAC 3080
Db 2928 GCTTCAAGCAACCACTTCTGAGTACAGACCCCTGAGGAGTATTTGATTTCAAC 2987
QY 3081 AGATTCACCTGCACTTTTCAACAGTGAAGTGGAGAGTCACTCAACCAATTTGGGGA 3140
Db 2988 AGATTCACCTGCACTTTTCAACAGTGAAGTGGAGAGTCACTCAACCAATTTGGGGA 3047
QY 3141 TTTGCGGCGCAAGAGACTCACTTCAACCTTCAACATCAAGTCAAGAGAGTCAAGAG 3200
Db 3048 TTTGCGGCGCAAGAGACTCACTTCAACCTTCAACATCAAGTCAAGAGAGTCAAGAG 3107
QY 3201 AATGATGGCTCAACCAACCACTTCAATTAACCTTCAACAGAGGCTTCAAGTCTTCTG 3260
Db 3108 AATGATGGCTCAACCAACCACTTCAATTAACCTTCAACAGAGGCTTCAAGTCTTCTG 3167
QY 3261 TCGAGTACCAAGCTTCCGTAAGTCTTGGCTTGGCGACCAAGGCTGCTCCCTCGCTC 3320
Db 3168 TCGAGTACCAAGCTTCCGTAAGTCTTGGCTTGGCGACCAAGGCTGCTCCCTCGCTC 3227
QY 3321 CCGGCGAGAGTTCATGATTTCCGCAATACGAGTCACTGAGCTCAACAGAGGCA 3380
Db 3228 CCGGCGAGAGTTCATGATTTCCGCAATACGAGTCACTGAGCTCAACAGAGGCA 3287
QY 3381 GCGGTGGAGAGTTCATCTTTTACATGCTGAGTGAATATTTCCCTTCTCAGATGCTGAGAG 3440

DB 3288 GCGGTGGAGCGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGCGAGATGTAAGACT 3347
QY 3441 GCGACAACTTTTACTGCTGACCTACCTTTGAGGAATGCTTCCACAGACGCTACGCG 3500
DB 3348 GGAATTAATCTTCAATTCACCTTACCTTGAAGATGATCTTTTCAACACAGCTACGCT 3407
QY 3501 CACAGCGAGCGCTGACCGGCTGATGATTCCTTCATCGACCAATACCTGTAATTAACCTG 3560
DB 3408 CACAGCGAGCGCTGATGATTCCTTCATCGACCAATACCTGTAATTAACCTG 3467
QY 3561 AACAGAACTCA---AAATCAGTCGGAAGTCCCAAAAACAAGACTTGTCTTTAAGCCGT 3617
DB 3468 AACAGAACTCA---AAATCAGTCGGAAGTCCCAAAAACAAGACTTGTCTTTAAGCCGT 3527
QY 3618 GGGGTTCGAGCTGACGATGCTGTTGACGCGCAAAACCTGCTGACCTGCTGTTATCGG 3677
DB 3528 GCTGGGCTGAGTCTATGCTTTGACAGCGCAAAATGCTGCTGAGCCCTGCTACCGG 3587
QY 3678 CACAGCGCGCTTCTTAAACAAAACAGACAAACAAACAGCAATTTTACCTGAGCTGCT 3737
DB 3588 CACAGCGAGCTTCTTAAACAAAACAGACAAACAAACAGCAATTTTACCTGAGCTGCT 3647
QY 3738 GCTTCAAAATATACTCAATGGCGCTGAATCCATCATCAACCTGCGACTGCTATGCGC 3797
DB 3648 GCGAGCAATATATCATCTCAATGGCGCGCATCGCTGCTGATCAACAGCAAGCTATGCGC 3707
QY 3798 TCACACAAAGACGACGAAAGCAAGTTTTCCTTCCATGAGCGGCTGATGTTTGGAAA 3857
DB 3708 AGTCACAAAGACGACGAAAGCAAGTTTTCCTTCCATGAGCGCAATCAATTAATTTGGCAA 3767
QY 3858 GAGAGCGCGCGAGCTTCAAACTGATGAGCAATGATGATGATGATGATGATGATGATGAT 3917
DB 3768 GAGAGCGCGCGAGCTTCAAACTGATGAGCAATGATGATGATGATGATGATGATGATGAT 3827
QY 3918 ATTAAAGCACTTAACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 3977
DB 3828 ATTGATGACCAATCTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 3887
QY 3978 AGCAGACGACCAATCTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 4037
DB 3888 AGCTCAAAATCACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 3947
QY 4038 GTGAGCAATGATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4097
DB 3948 GTGAGCAATGATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4007
QY 4098 GATGACACTTTTCACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 4157
DB 4008 GATGACACTTTTCACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGAT 4067
QY 4158 CAGATCTCATCAAAACAGCGCTGATGAGCAATGATGATGATGATGATGATGATGATGAT 4217
DB 4068 CAGATCTCATCAAAACAGCGCTGATGAGCAATGATGATGATGATGATGATGATGATGAT 4127
QY 4218 AAGTTGCTTCACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 4277
DB 4128 AAGTTGCTTCACTGATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 4187
QY 4278 GAGCTGAGAAAGAAACAGCAAGCGCTGATGAGCAATGATGATGATGATGATGATGATGAT 4337
DB 4188 GAGCTGAGAAAGAAACAGCAAGCGCTGATGAGCAATGATGATGATGATGATGATGATGAT 4247
QY 4338 GCAAAATCTGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4397
DB 4248 AACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4307
QY 4398 CCAATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4457
DB 4308 CCAATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4367
QY 4458 TGAATGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4517

DB 4368 TAATGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4427
QY 4518 ATAGCTTAC 4526
DB 4428 CTAGACTAC 4436
RESULT 11
US-10-205-942-2
; Sequence 2, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8151
; TYPE: DNA
; ORGANISM: Adeno-associated virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(8151)
; OTHER INFORMATION: AAHV/4 helper plasmid
US-10-205-942-2

Query Match 57.9%; Score 2730; DB 15; Length 8151;
Best Local Similarity 78.5%; Pred. No. 0;
Matches 3395; Conservative 0; Mismatches 890; Indels 41; Gaps 9;
QY 205 GTCTGATTAATGAGTGCAGTGAAGTCTTTGAGCAATTTTGGACACACAGCGGCATT 264
DB 121 GTCTGATTAATGAGTGCAGTGAAGTCTTTGAGCAATTTTGGACACACAGCGGCATT 179
QY 265 TAGGATTAATGAGTGCAGTGAAGTCTTTGAGCAATTTTGGACACACAGCGGCATT 323
DB 180 CTGGGATTAATGAGTGCAGTGAAGTCTTTGAGCAATTTTGGACACACAGCGGCATT 239
QY 324 GAGCAGACCATGCGGCTTCTGAGATGATGATCAAGTCCGAGCAGCTGAGC 383
DB 240 GCGCAGCCCGCATGCGGCTTCTGAGATGATGATCAAGTCCGAGCAGCTGAGC 299
QY 384 AGCAGCTGCGGCGCATTTGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 443
DB 300 GGCATGTCGCGCATTTGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 359
QY 444 TGCCCCCGGATTTCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 503
DB 360 TGCCCCCGGATTTCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
QY 504 AGAAGTGCAGGCGCATTTCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 563
DB 420 AGAAGTGCAGGCGCATTTCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 479
QY 564 TCTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623
DB 480 TTTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
QY 624 CCAAGGCGGATTTCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 683
DB 540 CCAAGGCGGATTTCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
QY 684 TGCAGACATCTACCGGCGGATGAGCGGACCTGCGCACTGCTTGGGAGCAAGCA 743
DB 600 TTGAGGATTTTACCGGCGGATGAGCGGACCTGCGCACTGCTTGGGAGCAAGCA 659
QY 744 CCGGTAATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 803

Db	660	CCAGAAATGCGCGGAGGCGGAAACAAAGTGTGTGAATGAGTCAATCCCAATTACT	719
QY	804	TCCTGCCCAAACTCAGCCCGCAGTGCAGTGGCCGTGTACCTAACATGAGAGATATATA	863
Db	720	TGCTCCCAAAACCGAGCTGAGCTCCAGTGGGCGGTGAACTATATGAAACGATTTTAA	779
QY	864	GCGCGCTGTTTAACTGTGCCGAGCGCAACCGCTCTGTGGCGCAGCACTTGACCCAGTCA	923
Db	780	GCGCCTGTTTGAATCTCAGCGGACCTTAAACGGTGTGTGGTGGCGAGCATCTGACGCACTGT	839
QY	924	GCCAGCCCAAGAGGAGAAACAAAGGAAATCTGSAACCCCAATTCTGAGGCGCTGTCAATCC	983
Db	840	CGCAGACCGAGAGAGAAACAAAGGAAATCGAAATCCCAATTCTATGTGGCCGGTGATCA	899
QY	984	GGTCAAAAACCTCCGCGGCTCAATAGAGCTGTGTGGGTGTGTGTGACCGGGGCATCA	1043
Db	900	GATCAAAAACCTTCAGCCAGGTACATGAGAGCTGTGTGGGTGTGTGTGACAAAGGGATTA	959
QY	1044	CTCTCGAAGACATGTGATCCAGAGAGACAGGCGCTCGTACATCTCTTCAACCGCGCTT	1103
Db	960	CCTCGGAACACAGTGATCCAGAGAGACAGGCGCTCATATCTCTTCATATGCGGCGCT	1019
QY	1104	CCAATCGCGGCTCCAGATTCAAAGCGCGCTGTGSAACAATGCGGGCAAGTCAATGGCGCTGA	1166
Db	1020	CCAATCTGCGGCTCCCAAAATCAAAGCTGCTTGGAAACAATGCGGAAAGATTTATGAGCTTGA	1079
QY	1164	CCAATTCGCGCGCCGACTACCTGTATGAGCCCGGCTCCGCGCGCGCACTTAAACCAAC	1222
Db	1080	CTAAACCGGCCCCGACCTACCTGTGTGGGCGACACCGCTGTAGAGACATTTTCCAGACATTC	1133
QY	1224	GCATCTACCGCATCTGTGAGCTGAACCGGCTACGAACCTTGCTACGCGGACTCGTCTTTC	1288
Db	1140	GGATTTATATAAATTTTGGAACTTAAACGGGATGAGTCCCAATATCGGCTTCCGTCTTTC	1199
QY	1284	TCGCTGTGGGCCAGAAAAGTTTCGGGAAAGCGCAACAATCTGTGTTTGGGCGGGCA	1343
Db	1200	TGGATGTGGCCACGAAAAAGTTTCGGCAAGAGAACCAATCTGTGTTTGGGCGGTGA	1255
QY	1344	CCACGGGCAAGACCAACATTCGCGGAAGCCATGCGCCAGCGCTGCTTCTACCGGCTCG	1403
Db	1260	CTACCGGAAACCAACATTCGCGGAGGCCATATGCCCACTGTGCTTCTACGGGTGG	1319
QY	1404	TCAACTGACCAATGAGAACTTTCCTTCAATGATTTGCGTGCACAAGATGTGATCTGTGT	1463
Db	1320	TAAACTGACCAATAGAACTTTCCTTCAACGATGTGTGCACAAGATGTGATCTGTGT	1375
QY	1464	GGAGAGAGGCAAGATGACCGGCAAGGTTGTGTGATTCGCGCAAGGCCATTTCTGGCGGCA	1522
Db	1380	GGAGAGAGGCAAGATGACCGGCAAGGTTGTGTGATTCGCGCAACCATTTCTCGAGGAA	1433
QY	1524	GCAAGGTGCGGTGACCCAAAGTGCATAGTCTGTGCGCCAGATTCGACCCACCCCGTGA	1583
Db	1440	GCAAGGTGCGGTGACCCAAAGTGCATAGTCTGTGCGCCAGATTCGACCCCGTGTGA	1499
QY	1584	TGCTCACTTCAACCAACATGTGCGCGGTGATTTGACGGGAAACGACCAACTTTCGAGC	1643
Db	1500	TCGTCACTTCAACCAACATGTGCGCGGTGATTTGACGGGAAACCAACGACTTTCGAAAC	1555
QY	1644	ACCAGACGCGTTTGCAGAGACCGGATGTTCAATTTTGAATCAACCGGCGGTCTGTGAGCATG	1700
Db	1560	ACCAGACGCGTTTGCAGAGACCGGATGTTCAATTTTGAATCAACCGGCGGTCTGTGATCATG	1619
QY	1704	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAAGATCTTTCGCGTGGGCGCAGATTCAG	1766
Db	1620	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAAGATCTTTCGCGTGGGCAAAAGATTCAG	1677
QY	1764	TGACCGAGGTGCGCATGATGTTCTATGTCAGAAAGGTGTGACCCAAACAAAGACCGGCGC	1822
Db	1680	TGCTTGAAGTGAAGATGATTTCTATGTCAAAAAGGTGTGACCCAAAGAAACACCGGCGC	1733
QY	1824	CCGATGACGCGGATTAACCGGACCCAACCGGGCGTGCCTCAAGTCGCGATTCATGTGA	1883

D	1740	CGA	GTGACGAGAAATTAATGAGAACCCAAAGGGGTGGCGAGTCA	GTGGCAGACAAATCGA	1799
Q	1884	CGT	CAGACGCGGAAGAGCTCCGGTGGACCTTTGCGCAGAGTAC	CAAAACAAATGTTCTC	1943
D	1800	CGT	CAGACGCGGGA---AGCTTCATCAACTACGACGACAGATGC	CAAAACAAATGTTCTC	1856
Q	1944	GT	CACCGGGGCATGCTTCAGATGCTGTTTCCCTGCAGAA	CATGCGAGAGAAATGATCAGA	2003
D	1857	GT	CACGCGGCGATGAATCTGATGCTGTTTCCCTGCAGAA	CAATGCGAGAGAAATGATCAGA	1916
Q	2004	ATT	CAACATTTTGCTTACGACGAGGAGCGAGAGCTGTTCA	GAGTGGTCTCCCGCGGTG	2063
D	1917	ATT	CAAAATATCTGCTTACCTACGACGAGACAGAAAGCTG	TTTAAGTGTCTTC---CGTGT	1973
Q	2064	CAG	AATTTCAACCGGT---CGTCAAGAAAGAGACGATTCG	GAATACCTTGTCCTCATTC	2120
D	1974	CAG	AATCTCAACCCGTTTCTGTGCTCAAAAAGGCGTATC	AGAAATCTGTCTCAATTCATC	2033
Q	2121	AT	CTGCTGGGGCGGGCTCCCGAAGTTGCTGCTCGGCT	CGACGATCTGGTCAACGTTGAGC	2180
D	2034	AT	ATACCAATGGGAAAGTGCACAGA---CGCTTGAC	TGCTGCGAGTCTGATCAATGAGATT	2090
Q	2181	TG	ATGATGATCTGTGTTTCTGACCAATTAATGACTTAA	ACAGATATGCTGCGATGTTAT	2240
D	2091	TG	ATGATGATCTGATCTTTGAAACAAATATGATTTAA	ATCAGATATGCTGCGATGTTAT	2150
Q	2241	CTT	CCAGATTTGGCTCGAGGACAACTCTCTGAGGGCAT	TCGACAGTGTGGACTTGAA	2300
D	2151	CTT	CCAAATTTGGCTCGAGGACAACTCTCTGAGGAA	TAAGACAGTGTGGAACTCMAA	2210
Q	2301	CT	CGAGACCCCGGAAGCCCAAGCCAAACACACAA	AAAGAGAGACGCGCGGGGCTGGTG	2360
D	2211	CT	CGGCGCACACACACCAAAAGCCCGACAGACGCG	CATTAAGACACACAGAGGGGCTTGTG	2270
Q	2361	CTT	CTCGGCTCAAGTACTCTCGACCCCTTCAACG	SACTCGACAGAGGGGAGCCGCTCAAC	2420
D	2271	CTT	CTCGGCTCAAGTACTCTCGACCCCTTCAACG	SACTCGACAGAGGGGAGCCGCTCAAC	2330
Q	2421	GC	GGCGGAGCGACGCGGCTCTCGAGACGACAG	AGGCTTACACACAGCACTCAAGCGGGT	2480
D	2331	GAG	GCACAGACGCGCGGCGCTCTCGAGACGACAG	AGGCTTACACACAGCACTCAAGCGGGT	2390
Q	2481	GAC	AAATCCGTAACCTCGCGGTATTAACCAAG	CCGACGCGGAGTTTACAGGCGCTTCAGAA	2540
D	2391	GAC	AAATCCGTAACCTCGCGGTATTAACCAAG	CCGACGCGGAGTTTACAGGCGCTTCAGAA	2450
Q	2541	GAT	ACGCTCTTTGGGGGCAACCTCGGCGAGAC	AGTCTTCCAGGCGCAAGAGCGGGTCTC	2600
D	2451	GAC	CAATGCTTTGGGGGCAACCTCGGCGAGAC	AGTCTTCCAGGCGCAAGAGCGGGTCTC	2510
Q	2601	GAA	CCCTCGGCTCTGTTGTTGAGAAAGGCGCT	TACAGCGGCTCTCGGAAAGAAACGTCCGGTA	2660
D	2511	GAA	CCCTCTTGCGTCTGTTGTTGAGAAAGGCGCT	TACAGCGGCTCTCGGAAAGAAAGACCGTTG	2570
Q	2661	GAG	ACATGTCGCACAAAGCCAGACTCTCTCGG	GCATTCGGGACATAGAGCCGACAGACCC	2720
D	2571	ATT	GAATATCCCCCGACAGCCCGACCTCTCA	CGGGTATGAGCAAAAAGGCGACAGACCCG	2630
Q	2721	GCT	AAAAAGACATCAATTTTGGTCAATCTG	GCACACTCAGAGTCACTCCCGATCCACAA	2780
D	2631	GCT	AAAAAGACATCAATTTTGGTCAATCTG	GCACACTCAGAGTCACTCCCGATCCACAA	2681
Q	2781	CT	CTCGAGAGACCTTCAGACAAACCCCGCT	GCTGTGGAACTCTACATGATGCTTACGC	2840
D	2682	CT	CTAGAGGATCAATTCGAGAGCCATGCTGATG	-----ACAGTGAATGCGTGACGACA	2735
Q	2841	GCT	GGCGCACCAATGCGACATATACAAAGG	CGCCGACGAGTGGGTAAATGCTTCA	2900
D	2736	GCT	GGCGCACCTGACATGCGAGGCGGACAAAGT	GCCATGAGAGTGGGTAAATGCTTCCGGT	2795
Q	2901	AA	TTTGCAATTTGCAATTCACATGCTGGG	GCAGACATCAACACACAGACCCGCAAC	2960
D	2796	GAT	TGGCATTTGCAATTCACCTGCTGAGG	GCACAGTCACTCAACACAGACCCGCAAC	2855

QY 2961 TGGGCTTGGCCACTTCAATTAACCACTCTGACAGGAATCTCAGTGTTCACGCGG 3020
 DB 2856 TGGGTCTTGGCCACTTCAATTAACCACTCTGACAGGAATCTCAGTGTTCACGCGG 2903
 QY 3021 GCCAGCAAGCAACCACTTCTGAGTACAGCACTCCCTGGGGATTTTGAATTTCAAC 3080
 DB 2904 AGCTGACATGCAACCACTTCAAGGATTTCCACCTCCCTGGGGATTTTGAATTTCAAC 2963
 QY 3081 AGATTCACATGCACTTTTCAACCACTGATGAGGAGTCAATCAACCAATTTGGGGA 3140
 DB 2964 CGCTTCACATGCACTTTTCAACCACTGATGAGGAGTCAATCAACCAATTTGGGGA 3023
 QY 3141 TTCGCGGCAAGAGACTCAATTCCTTCAACCACTGATGAGGAGTCAACGAG 3200
 DB 3024 ATGCGACCAAGAGACTGCGGATCAATTTCAACCACTGATGAGGAGTCAACGAG 3083
 QY 3201 AATGATGCGTCAACCACTGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3260
 DB 3084 TCGAAGCGGAGCAACGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3143
 QY 3261 TCGGAGTCAAGCTTCCGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3320
 DB 3144 TCGTTCGATCAAGCTTCCGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3203
 QY 3321 CCGGCGGAGCTGATGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3380
 DB 3204 CAGGACAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3263
 QY 3381 GCGGTGAGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3440
 DB 3264 GCGGTGAGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3323
 QY 3441 GCGGTGAGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3500
 DB 3324 GCGGTGAGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3383
 QY 3501 CAGGCGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3560
 DB 3384 CAGGCGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3443
 QY 3561 AACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAGAACTTGTGTTAGCCGTGG 3620
 DB 3444 AGAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAGAACTTGTGTTAGCCGTGG 3503
 QY 3621 TCTTCAGCTGAGTCTGTTGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3680
 DB 3504 GAGGCGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3563
 QY 3681 CAGGCGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3740
 DB 3564 CAGGCGGAGCTTCTTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3623
 QY 3741 TCAAAATATTAATCTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3800
 DB 3624 ACCAAGTACCACTCAATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3683
 QY 3801 CACAAAGACAGAACTGTTTCTTCCATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3860
 DB 3684 CACAAAGACAGAACTGTTTCTTCCATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3743
 QY 3861 AGGCGGAGCTTCAAACTGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3920
 DB 3744 GCGTTCAGGAGAACTGATGAGGAGTCAACCACTGATGAGGAGTCAACGAG 3803
 QY 3921 AAGGCACTTAACCTGTGAGCAACGAAATTTGGGACCGTGGAGTCAATTTTCAGAGC 3980
 DB 3804 GGAACAACCAATCCGTTGCTAGGAGCAAGATGTTCTGATCTACCACTTCAGAGA 3863
 QY 3981 AGGAGCAACAACCTGTGAGCAAGATGTTCTGATCTACCACTTCAGAGA 4040
 DB 3864 GGAACAACCAAGTCTACCGAGATGTTCTGATCTACCACTTCAGAGA 3923

QY 4041 TGGCAAGATAGAGACCTGATCTGAGGAGTCCATTTGGGCAAAATTCCTCAGAGAT 4100
 DB 3924 TGGCAGAGAGAGATGATCTTCAAGGAGGAGTCCATTTGGGCAAAATTCCTCAGAGAT 3983
 QY 4101 GGAACCTTCAACCTGTTCTCTTATGAGGAGGCTTTGAGTCAAGAACCCGCTCTCAG 4160
 DB 3984 GGAACCTTCAACCTGTTCTCTCTCAGGAGGATGAGTCAAGAACCCGCTCTCAG 4043
 QY 4161 ATCTGATCAAAACAGAGCTGTTCTGAGGAGTCCGAGGAGTTCAGTCAAG 4220
 DB 4044 ATCTGATCAAAACAGAGCTGTTCTGAGGAGTCCGAGGAGTTCAGTCAAG 4103
 QY 4221 TTGCTTCAATTCACCACTTCAAGTCAAGAGTCAAGTCAAGTCAAGTCAAG 4280
 DB 4104 TTGCTTCTTCAATTCACCACTTCAAGTCAAGAGTCAAGTCAAGTCAAGTCAAG 4163
 QY 4281 CTGCAAG 4340
 DB 4164 CTGCAAG 4223
 QY 4341 AATTCGCAAGCTGATTTTATGAGCAACAATGAGTCAATGAGTCAAGTCAAG 4400
 DB 4224 AATTCGCAAGCTGATTTTATGAGCAACAATGAGTCAATGAGTCAAGTCAAG 4283
 QY 4401 ATTGCAAGCTGATTTTATGAGCAACAATGAGTCAATGAGTCAAGTCAAG 4460
 DB 4284 ATTGCAAGCTGATTTTATGAGCAACAATGAGTCAATGAGTCAAGTCAAG 4343
 QY 4461 TTGCTTCAATTCACCACTTCAAGTCAAGAGTCAAGTCAAGTCAAGTCAAG 4520
 DB 4344 TTGCTTCAATTCACCACTTCAAGTCAAGAGTCAAGTCAAGTCAAGTCAAG 4403
 QY 4521 GCTTAC 4526
 DB 4404 GACTAC 4409

RESULT 12
 US-10-205-942-1/c
 ; Sequence 1, Application US/10205942
 ; Publication No. US20030053990A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
 ; APPLICANT: Rabinowitz, Joseph E.
 ; APPLICANT: Samulski, Richard J.
 ; APPLICANT: Xiao, Weidong
 ; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
 ; FILE REFERENCE: 5470-186
 ; CURRENT APPLICATION NUMBER: US/10/205,942
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 7214
 ; TYPE: DNA
 ; ORGANISM: Adeno-associated virus
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(7214)
 ; OTHER INFORMATION: AAV2/4 helper plasmid
 US-10-205-942-1

Query Match 51.8%; Score 2446.2; DB 15; Length 7214;
 Best Local Similarity 75.1%; Pred. No. 0;
 Matches 3277; Conservative 0; Mismatches 1008; Indels 76; Gaps 15;
 QY 205 GTCTGATTAAGGCTGACGAGTCTTGTGAGCATTTTGGAGACCAAGTGGCATT 264
 DB 5001 GTCTGATTAAGGCTGACGAGTCTTGTGAGCATTTTGGAGACCAAGTGGCATT 4943
 QY 265 TAGGTAATATGAGCGGAGTGAAGAGAGATCTCCATTTTG-ACCGGAAATTTGAAC 323
 DB 4942 CTGGTAATTAAGCGGAGTGAAGAGAGATCTCCATTTTGAAACGGAGGTTGAAC 4883

QY 324 GAGGAGGAGCCATGCGCGCTTCTACGAGATGATGCAAGGTGCGGAGCGACCTGAGCG 383
DB 4882 GCGAGCGCGCATGCGCGGTTTTCAGAGATTGATTAAGGTCCCGACCACTTGAACG 4823
QY 384 AGCACTGCGCGGAGATTTCTGACTGTTTGTAGCTGGGTGGCGAGAGAGATGAGAGC 443
DB 4822 GGCATCTGCGCGGAGATTTCTGACAGCTTTGTGAATGGGTGGCGAGAGAGATGAGAGT 4763
QY 444 TGCCCCCGGAGATTCTGACATGATCTGAATCTGAATTGAGCAGGACCCCTGACCGTGGCG 503
DB 4762 TGCGCGGAGATTCTGACATGATCTGAATCTGAATGAGCAGGACCCCTGACCGTGGCG 4703
QY 504 AGAAGCTGACGCGGAGCTTCTGCTGCAATGCGCGCGGTGAATGAAGGCCCGCGAGGCC 563
DB 4702 AGAAGCTGACGCGGAGCTTCTGCTGCAATGCGCGGTGAATGAAGGCCCGCGAGGCC 4643
QY 564 TCTTCTTGTTCAGTTTGAGAAAGGCGAGCTCTACCTTCACCTCCATATTTCTGTGAGA 623
DB 4642 TTTTCTTGTTCAGTTTGAGAAAGGAGAGCTACTTCACATGCAAGTGTCTGTGAAA 4583
QY 624 CCAGCGGGGTCAATTCATGTGCTGGCGCGCTTCTGAGTCAAGTTAGGAGCAAGCTGG 683
DB 4582 CCAGCGGGGTCAATTCATGTGCTGGCGCGCTTCTGAGTCAAGTTAGGAGCAAGCTGA 4523
QY 684 TGCAAGCATCTACCGCGGGAGTCGAGCGGACCTGCGCCAACTGCTGGCGGTGACAGA 743
DB 4522 TTTCAAGAAATTTACCGCGGGATCGAGCGGACTTTGCAAACTGCTTGGCGGTGACAAA 4463
QY 744 CGCGTAATGCGCGGAGAGGGGAGAAACAAGTGTGAGCAGATGCTCAATCCCACTACC 803
DB 4462 CCAGAAATGCGCGGAGAGGGGAGAAACAAGTGTGAGTGAATGCTCAATCCCACTACT 4403
QY 804 TCTTGTCCCAAGATCTACGCGCGAGCTGCGAGTGGGTGAGCTAAATGAGAGATTAATA 863
DB 4402 TGCTCCCAAAACCGAGCTGAGCTCAAGTGGGTGAGCTAAATGAGAGATTAATA 4343
QY 864 GCGCTGTTTGAACCTGGCGCGAGCGCAAAACGCTCGTGGCGAGACCTGACCCACGCTCA 923
DB 4342 GCGCTGTTTGAATCTACGAGAGCGTAAACGTTGTGTGGCCAGATCTGACGACGTGT 4283
QY 924 GCGAGACCTGAGAGCAGAAACAAGAGATCTGAATCCCAATTTGACGCGCTGTCACTC 983
DB 4282 CGCAGACGAGAGCAGAAACAAGAGATCTGAATCCCAATTTGATGCGCGGTGATCA 4223
QY 984 GGTCAAAACCTCCGCGCGCTACATGAGCTGTGCGGTGGCTGTGAGCCGGGAGATCA 1043
DB 4222 GATCAAAACCTTCAGCCAGGTACATGAGCTGTGCGGTGGCTGTGAGCAAGGGAGATTA 4163
QY 1044 CCTCCGAGAGCAGTGTGATCCAGAGAGCAGGCTCGTACATCTCTTCAACCGCTT 1103
DB 4162 CCTCGAGAGAGCAGTGTGATCCAGAGAGCAGGCTCGTACATCTCTTCAATCGGCT 4103
QY 1104 CCAACTGCGGCTCCAGATCAAGCGCTCTGACATGCGCGCAAGTCAATGCGCTGA 1163
DB 4102 CCAACTGCGGCTCCCAATCAAGCGCTCTGACATGCGCGCAAGTCAATGAGCTGA 4043
QY 1164 CCAATTCGCGCGCGCTACCTGTGTAGGCCCCGCTCGCGCGCGAGCATTAACCAACC 1223
DB 4042 CTAAACCGCGCGCGCTACCTGTGTAGGCGCGAGCGCGTGAAGACATTTCCAGCAATC 3983
QY 1224 GCATCTACCGCATCTGAGAGCTGAACCGCTACGACTGCTACGCGCGCTCGCTTTTC 1283
DB 3982 GGAATTTAATAATTTTGAATCTAAACGGGTACGATCCCAATTAAGCGCTTCGCTTTTC 3923
QY 1284 TGCGCTGGCGCGCAAAAGTTTGGGAAGCGCAACACATCTGCTGTTTGGCGCGCA 1343
DB 3922 TGCGATGGCGCGCAAAAGTTTGGGAAGCAACACATCTGCTGTTTGGCGCGCA 3863
QY 1344 CCAGCGGAGAGACCAATCTGCGGAGAGCATGCGCGCGCTTCTCAACGCTGGC 1403
DB 3862 CTACCGGAGAGACCAATCTGCGGAGAGCATGCGCGCGCTTCTCAACGCTGGC 3803

QY 1404 TCACTGAGCAATGAGAACTTCTCTCAATGATTGCGTGAAGAGTGTGATGTGT 1463
DB 3802 TAAACTGAGCAATGAGAACTTCTCTCAATGATTGCGTGAAGAGTGTGATGTGT 3743
QY 1464 GGGAGAGGGCAATGAGCGGCCAAGTCTGAGTCTCCGCAAGGCCATTTCTGGCGCA 1523
DB 3742 GGGAGAGGGCAATGAGCGGCCAAGTCTGAGTCTCCGCAAGGCCATTTCTGGAGAA 3683
QY 1524 GCAAGGTGCGCGTGAACCAAAAGTGAAGTCCGCGCGAGTCAAGCCCGCGTGA 1583
DB 3682 GCAAGGTGCGCGTGAACCAAAAGTGAAGTCCGCGCGAGTGAAGCCGATCCGCTGA 3623
QY 1584 TCGTCACTTCAACCAACCATGTCGCGCGTGAATGAGGAAACAGACCACTTGAAGC 1643
DB 3622 TCGTCACTTCAACCAACCATGTCGCGCGTGAATGAGGAACTCAACCACTTGAAC 3563
QY 1644 ACCAGACCGCTTGAAGAGCCGAGATTTCAATTTGAATCTACCCGCGCTGAGAGATG 1703
DB 3562 ACCAGACCGCTTGAAGAGCCGAGATTTCAATTTGAATCTACCCGCGCTGAGATG 3503
QY 1704 ACTTTGGCAAGTGAACAAAGCAGAAAGTCAAAAGTCTTCCGCTGGCGCAGGATCAG 1763
DB 3502 ACTTTGGCAAGTGAACAAAGCAGAAAGTCAAAAGTCTTCCGCTGGCGCAGGATCAG 3443
QY 1764 TGAACGAGTGGCGCATGAGTCTTACGTCAGAAAGGTGAGGCCAACAAAGACCCGCC 1823
DB 3442 TGGTGAGGTGAGAGCATGATTTCTAGTCAAAAGGTGAGGCCAAAGAAAGCCGCC 3383
QY 1824 CCGATGACGCGGATTAAGAGCAGCCAGCGGCGCTGCGCTCAATGCTGCGGATCTCA 1883
DB 3382 CCGATGACGCGGATTAAGAGCAGCCAGCGGCGCTGCGGATGAGTGCACACCATCA 3323
QY 1884 CGTCAAGCGGAGAGAGCTCCGCTGAGCTTTCGCGCAGAGTCAAAACAAATGTTCTC 1943
DB 3322 CGTCAAGCGGAGAGAGCTCCGCTGAGCTTTCGCGCAGAGTCAAAACAAATGTTCTC 3286
QY 1944 GTCAAGCGGAGATGCTTCAAGATCTGTTTCTGCAAGATGCAATGCAATGATCA 2003
DB 3286 GTCAAGCGGAGATGCTTCAAGATCTGTTTCTGCAAGATGCAATGCAATGATCA 3206
QY 2004 ATTTCAATTTGCTTCAAGCAGCAGAGATGTTTGAAGTCTTCCCGCGGTGT 2063
DB 3205 ATTTGACATTTGCTTCAAGCAGCAGAGTCAAGATGTTTGAAGTCTTCCCGGTGT 3149
QY 2064 CAGATCTCAAC-----CGGTGTCAGAAAGAGAGTTCGAAACTGTGCCATTC 2117
DB 3148 CAGATCTCAACCGGTGTCTGTGTCTGAGAGAGAGAGTTCGAAACTGTGCCATTC 3089
QY 2118 ATCATCTGAGGAGCGGCTCCGAGATGCTTGTGCGGCTGCGATCTGTCACGTTG 2177
DB 3088 ATCATCATGAGGAGCGGCTCCGAGATGCTTGTGCGGCTGCGATCTGTCACGTTG 3029
QY 2178 ACTGTGATGCTGTGTTTCTGAGCATTAATGATTAACACAGTATGCTGCGCATGT 2237
DB 3028 ACTGTGATGCTGTGATGAGTAAATTAATGATTAACACAGTATGATCTGCGCATGT 2297
QY 2238 TATCTTCAATTTGCTGAGAGCAACCTCTGAGAGGCTTGTGCGAGTGTGAGGCTT 2297
DB 2971 TATCTTCAATTTGCTGAGAGCAACCTCTGAGAGGCTTGTGAGAGTGTGAGGCTT 2297
QY 2298 AAACCTGAGACCCGAGAGCCCAAGCCCAACAGCAAAAGCAGAGCAGCGCGGTCTG 2357
DB 2911 CAACCTGAGACCCGAGAGCCCAAGCCCAACAGCAAAAGCAGAGCAGCGCGGTCTT 2852
QY 2358 GTGCTTCTGCTTCAAGTACTCTGAGACCTTCAAGGATCTGCAAGAGGAGGCCGCTC 2417
DB 2851 GTGCTTCTGCTTCAAGTACTCTGAGACCTTCAAGGATCTGCAAGAGGAGGCCGCTC 2792
QY 2418 AACGCGGAGAGCAGCGGCTCTGAGAGCAGCAAGAGCCTTCAAGCAGCAGCTCAAGCG 2477
DB 2791 AACGCGGAGAGCAGCGGCTCTGAGAGCAGCAAGAGCCTTCAAGCAGCAGCTCAAGCG 2732
QY 2478 GGTGACATCTGATCTGCGGTATTAACCAAGCGCGAGGTTTCAAGAGCGCTTGA 2537

Db 2731 GGTGACAAACCCCTCACTCAAGTACAAACGCGGAGGTTCCAGCGGCTTCAG 2672
Qy 2538 GAAGATACGCTTTTGGGGGCAACCTGGGGAGACGCTTCCAGGCGCAAGAGCGGTT 2597
Db 2671 GGGGACACATCGTTTGGGGGCAACCTGGGAGACGCTTCCAGGCGCAAGAGCGGTT 2612
Qy 2598 CTCGAACCTCTGGGTCTGTGTGAGGAGCGCTAAGACGCGCTCTGGAAAGAAAGCTCCG 2657
Db 2611 CTGGAACCTCTTGGTCTGTGTGAGGAGCGGCTAAGACGCGCTCTGGAAAGAAAGCG 2552
Qy 2658 GTAGAGACGTGCGCAAGAGCGGAGCTCTCTGGGCAATCGGCAAGACGCGCAG 2717
Db 2551 TTGATTAATCCCGCCAGAGCGCGGCTCTCCAGCGGCTATCGCAAAAAGGCAAGCAG 2492
Qy 2718 CCGGCTAAAAAGATCTCAATTTTGTGCAAGTGGCACTCAGAGTACGCTCCGATCCA 2777
Db 2491 CCGGCTAAAAAGATCTGTTTTT-----CGAAGACGAAACTGAGAGCGGCAAGCGA 2441
Qy 2778 CAACCTCTGGAGAACTCCAGCAACCGCGCTGTGGGAACTCTAATAGGCTTCA 2837
Db 2440 CCGGCTTGAAGGATTAATCTTCGAGGCGCATGTCTGATG-----ACAGTGAAGTGCCTGCA 2387
Qy 2838 GGGCGGTGGCGCAACCAATGGAGCAATTAAGAGCGCGCGAGTGGGTAATGCTTCA 2897
Db 2386 GCAAGTGGCGGAGTGAAGTGAAGGCGGCAAGAGTGCAGTGAAGTGGGTAATGCTTCCG 2327
Qy 2898 GGAATTTGGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 2957
Db 2326 GGGATTTGGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 2267
Qy 2958 ACCGCGGCTCTGGGCACTCAATTAACCACTCTCAAGAGAAATCTCAAGTGTCAAG 3017
Db 2266 ACCGCGGCTCTGGGCACTCAATTAACCACTCTCAAGAGAAATCTCAAGTGTCAAG 2219
Qy 3018 GGGGCGGCAAGCAACCACTCTCTGAGCAAGCAACCGCGGCTTCTTGAATTC 3077
Db 2218 GAGAGCTGAGTCAACCACTCTCAAGAGAAATCTCAAGTGTCAAG 2159
Qy 3078 AACAGATTCACATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 3137
Db 2158 AACCGCTTCACTGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 2099
Qy 3138 GGAATTCGCGGCGGCAAGAGACTCACTCAATTCGCAATTCGCAATTCGCAATTCGCA 3197
Db 2098 GGAATTCGCGGCGGCAAGAGACTCACTCAATTCGCAATTCGCAATTCGCAATTCGCA 2039
Qy 3198 AACGATGATGCGGCTCAACCACTCTCAATTCGCAATTCGCAATTCGCAATTCGCA 3257
Db 2038 AACGATGATGCGGCTCAACCACTCTCAATTCGCAATTCGCAATTCGCAATTCGCA 1979
Qy 3258 GACTCGGAGTACCAAGCTTCCGTAAGTCTGCGCTGCGCAACGAGGCTGCTCCG 3317
Db 1978 GACTCGGAGTACCAAGCTTCCGTAAGTCTGCGCTGCGCAACGAGGCTGCTCCG 1919
Qy 3318 TTCCCGGCGGAGCTTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 3368
Db 1918 TTCCCGGCGGAGCTTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 1859
Qy 3369 AATGGAGCGGCAACCGGCTGAGCTTCAATTCGCAATTCGCAATTCGCAATTCGCA 3428
Db 1858 ACTTGGAGCAACGAGCTGAGCAAGAAAGCTTCAATTCGCAATTCGCAATTCGCA 1799
Qy 3429 ATGCTGAGAAAGGCGCAACCTTCACTTCAAGTCAACCTTGAAGAGTGCCTTTCAC 3488
Db 1798 ATGCTGAGAAAGGCGCAACCTTCACTTCAAGTCAACCTTGAAGAGTGCCTTTCAC 1739
Qy 3489 AGCAGCTAGCGGCAAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 3548
Db 1738 TCGATGAGCGGCAAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 1679
Qy 3549 CTGTATTAATCTGAA---CAGAACTCAAAATCAAGTCCGGAAGTCCCAAAACAGAGCTTG 3605

Db 1678 CTGTGGGAGTGCATTCAGACCAACCGGAAACCAACCTGATGCGGAGCTGCACACC 1619
Qy 3606 CTGTGTTAGCGGTGGGTCTCCAGTGGGAGTCTGTTCAGGCCAAATCTGGCTTGA 3665
Db 1618 AACTTAAACCAAGCTGGGCGCTTCAACTTTTCAACTTTTAAAAAGAGCTGGCTCGG 1559
Qy 3666 CCTGTTATCGGAGAGCGGCTTCTTCAAAACAAACAGACAAACAGCAATTTT 3725
Db 1558 CTTTCAATCAAGAGAGGCTTCTTCAAGAGCTGCCAATCAAAATCAAGATCCCTGCC 1499
Qy 3726 ACCTGAGCTGGTCTTCAAAATTAAC-----CTCAATGGGCGTGAATCC 3770
Db 1498 ACCGGGTCAAGACAGTCTCAATCAATTAAGAGAGCGACAGCACTGTGAGCGGAATGAGT 1439
Qy 3771 ATCATCAACCTTGGCACTGTATGCTCTCAACAAAGAGAGCAAGCAAGTCTTCC 3830
Db 1438 GCGCTGACCGGCGGAGCTTCAATGAGCGAGCTGAGCTGCGGAGCAAGATTC---AGC 1382
Qy 3831 ATGAGCGGTGTCAATGATTTTGGAAAGAGAGCGGAGCTTCAAACTGCAATGGAC 3890
Db 1381 AACGCGAGCTCATCTTGGGCGGCTTAAACAGAGCGCAACGCGCACCGTACCGG 1322
Qy 3891 AATGCTATGATTAACAGAGAGAAATTAAGCACTTAAACCTTGGGCAACGAAAG 3950
Db 1321 ACTCTGATCTTCACTGTGAGAGAGCTGGAGCGCAACGCAACCGATTAAGCATG 1262
Qy 3951 TTGGGAGCGGTGCAATTTTCCAGAGAGAGAGCAAGACCTTGGAGCGGAGATG 4010
Db 1261 TGGGAGCACTTACCTGGCGGTGACAGAGCAACAGCAACCTGCGACCGTGGACAGACTG 1202
Qy 4011 CATGCTATGGAGCACTTACCTGGCATGTGTGGCAAGATTAAGAGCGTGAACCTGAGG 4070
Db 1201 ACAGCTTGGAGCGGCTGCTGAGATGTCTGCAAAACAGAGCAATTAACAGG 1142
Qy 4071 CCCATTTGGGCGCAAAATCTCTCAACAGATGAGCACTTTCACCGCTCTCTTATGAGG 4130
Db 1141 CCCATTTGGGCGCAAAATCTCTCAACAGATGAGCACTTTCACCGCTCTCTTATGAGG 1082
Qy 4131 GCGTTGAGCTCAAGAACCGGCTCTCTGATCTCTATCAAAACAGCGCTGTCTGG 4190
Db 1081 GCGTTGAGCTCAAGAACCGGCTCTCTGATCTCTATCAAAACAGCGCGTACCTGG 1022
Qy 4191 AATCTCCGCGGAGTTTCACTCAACAGATGAGCACTTTCACCGCTCTCTTATGAGG 4250
Db 1021 AATCTCCGAGCAACCTTCACTCACTTACCTCGGTAATCTCTTATTAAGAGACT 962
Qy 4251 GGAAGATGAGTGAAGAAATGAATGAGAGCTGCAAGAAAGAAACAGCAAGCGCTGAAT 4310
Db 961 GGCAGAGTGCAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 902
Qy 4311 CCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4370
Db 901 CCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 842
Qy 4371 AACAAATGACTTATCTGAGCGTCCCAATTTGAGACCGTATCTTACCGTCCCGT 4430
Db 841 GCGGCTGGAAATTAACCTGAGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 782
Qy 4431 TAATTAAGTGAATTAACCGGCTGATGCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4489
Db 781 TAATTAAGTGAATTAACCGGCTGATGCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 722
Qy 4490 CCGTCTTATCTTATC---GGTTACAGTGTATTAAGTGAAGTGAAGTGAAGTGAAG 4529
Db 721 CCGTCTTATCTTATC---GGTTACAGTGTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 681

RESULT 13
US-09-792-630-12
; Sequence 12, Application US/09792630
; Patent No. US20020168640A1
; GENERAL INFORMATION:
; APPLICANT: Lt, Min

APPLICANT: Dahiya, Basail I.
TITLE OF INVENTION: BIOCHIPS COMPRISING NUCLEIC ACID/PROTEIN CONJUGATES
FILE REFERENCE: A-70295/RFT/RMS/RMK
CURRENT APPLICATION NUMBER: US/09/792,630
CURRENT FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 1872
TYPE: DNA
ORGANISM: adeno-associated virus 1
US-09-792-630-12

Query Match 39.7%; Score 1872; DB 11; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

335 ATGCGGGGCTTCTACGAGATGATGATCAAGGTGCGGAGCCAGCTGAGCAGACCTGCG 394
1 ATGCGGGGCTTCTACGAGATGATGATCAAGGTGCGGAGCCAGCTGAGCAGACCTGCG 60
395 GGCATTTCTGACTGTTTGTGAGCTGGTGGCCGAGAGAAATGGAGCTGCCCGGAT 454
61 GGCATTTCTGACTGTTTGTGAGCTGGTGGCCGAGAGAAATGGAGCTGCCCGGAT 120
455 TCTGACATGATCTGAATCTGATTGAGAGGACCCCTGACCTGAGCCGAGAGCTGAG 514
121 TCTGACATGATCTGAATCTGATTGAGAGGACCCCTGACCTGAGCCGAGAGCTGAG 180
515 GCGACTTCTGATCTCAATGAGCGCGGCTGATGATGAGGCGCGGAGCCCTCTTTGTT 574
181 GCGACTTCTGATCTCAATGAGCGCGGCTGATGATGAGGCGCGGAGCCCTCTTTGTT 240
575 CAGTTCCAGAAAGGCGAGCTCTTCACTTCACTTCAATTTCTGTGAGAGACCAAGGGGTC 634
241 CAGTTCCAGAAAGGCGAGCTCTTCACTTCACTTCAATTTCTGTGAGAGACCAAGGGGTC 300
635 AATTCATGCTGCTGGGCGGCTTCTGATGATGATGAGGAGCAAGCTGTGAGAGCCTATC 694
301 AATTCATGCTGCTGGGCGGCTTCTGATGATGATGAGGAGCAAGCTGTGAGAGCCTATC 360
695 TACCGCGGATGAGCGGAGCCCTGAGCTGATGAGGAGCAAGCTGTGAGAGCCTATGTC 754
361 TACCGCGGATGAGCGGAGCCCTGAGCTGATGAGGAGCAAGCTGTGAGAGCCTATGTC 420
755 GCGGAGGCGGAGCAAGCTGTGAGAGCTGATGATGAGGAGCAAGCTGTGAGAGCCTATGTC 814
421 GCGGAGGCGGAGCAAGCTGTGAGAGCTGATGATGAGGAGCAAGCTGTGAGAGCCTATGTC 480
815 ACTCAGCCGAGCTGAGTGGGCGTGAAGCTAACTGAGAGAGTATTAAGAGCGCTGTTTG 874
481 ACTCAGCCGAGCTGAGTGGGCGTGAAGCTAACTGAGAGAGTATTAAGAGCGCTGTTTG 540
875 AACTGCGGAGCGCAAAAGGCTGTGAGCGAGAGCCTGACCTGACGACGAGACCCAG 934
541 AACTGCGGAGCGCAAAAGGCTGTGAGCGAGAGCCTGACCTGACGACGAGACCCAG 600
935 GAGGAGAAAGAGAGATCTGAACCCCAATTTCTGACGCGCTGTGATCCGGTCAAAAAC 994
601 GAGGAGAAAGAGAGATCTGAACCCCAATTTCTGACGCGCTGTGATCCGGTCAAAAAC 660
995 TCCGCGGCTCATGAGAGCTGTGAGGCTGTGAGAGCGAGAGCAATCACTCCGAGAG 1054
661 TCCGCGGCTCATGAGAGCTGTGAGGCTGTGAGAGCGAGAGCAATCACTCCGAGAG 720
1055 CAGTGATCCAGAGAGAGCAGGCTGTGATCTTCTTCAAGCGCGCTTCAACTCGGG 1114
721 CAGTGATCCAGAGAGAGCAGGCTGTGATCTTCTTCAAGCGCGCTTCAACTCGGG 780
1115 TCCGAGATCAAGGCGCTGTGAGCAATGCGGAGAGATCAATGAGGCTGACCAATTCGGG 1174
781 TCCGAGATCAAGGCGCTGTGAGCAATGCGGAGAGATCAATGAGGCTGACCAATTCGGG 840

1175 CCGACTACTGTAGAGCCCGCTCCGCGGAGCATTAAACCAACGCAATCAACG 1234
841 CCGACTACTGTAGAGCCCGCTCCGCGGAGCATTAAACCAACGCAATCAACG 900
1235 ATCTGAGCTGAAGGCTTACGAACTGCTTACGCGGCTCGCTCTTCTCGGCTGGGC 1294
901 ATCTGAGCTGAAGGCTTACGAACTGCTTACGCGGCTCGCTCTTCTCGGCTGGGC 960
1295 CAGAAAGGTTGGGAGGCGCAACATCTGCTGTTGGGCGGAGCAACGAGGCAAG 1354
961 CAGAAAGGTTGGGAGGCGCAACATCTGCTGTTGGGCGGAGCAACGAGGCAAG 1020
1355 ACCAATATCGGAGAGCAGTCCGAGCGGCTGCTTCTACGCTGCTGCACTGAGAC 1414
1021 ACCAATATCGGAGAGCAGTCCGAGCGGCTGCTTCTACGCTGCTGCACTGAGAC 1080
1415 AATGAGACTTTCCTTCAATGATTTGCTGAGCAAGATGATCTGTGTGGAGAGGGC 1474
1081 AATGAGACTTTCCTTCAATGATTTGCTGAGCAAGATGATCTGTGTGGAGAGGGC 1140
1475 AAGATGAGGCGCAAGGCTGTGAGTCCGAGGCGCATTTCTGCGGCGGAGCAAGTGGC 1534
1141 AAGATGAGGCGCAAGGCTGTGAGTCCGAGGCGCATTTCTGCGGCGGAGCAAGTGGC 1200
1535 GTGAGCCAAAGTGAAGTGTGCTCGGCGAGATGAGCCCAACCCCGTGTGCTACCTCC 1594
1201 GTGAGCCAAAGTGAAGTGTGCTCGGCGAGATGAGCCCAACCCCGTGTGCTACCTCC 1260
1595 AACACCAACTGTGCGCGCTGATTTGACGAGGAGAACAGCAACCTTTCAGACCCAGCCG 1654
1261 AACACCAACTGTGCGCGCTGATTTGACGAGGAGAACAGCAACCTTTCAGACCCAGCCG 1320
1655 TTGACGAGCCGAGATGTTCAATTTGAACTCACCGCGCTGTGAGCATGCTTTGCAAG 1714
1321 TTGACGAGCCGAGATGTTCAATTTGAACTCACCGCGCTGTGAGCATGCTTTGCAAG 1380
1715 GTGACAAAGCAGAGAGTCAAGAGTCTTCCGCTGAGGCGGAGATCACTGACCGAGTG 1774
1381 GTGACAAAGCAGAGAGTCAAGAGTCTTCCGCTGAGGCGGAGATCACTGACCGAGTG 1440
1775 GCGCATGAGTCTTACGTCAGAAAGGCTGAGCCCAAAAGACCCCGCGATGACGCG 1834
1441 GCGCATGAGTCTTACGTCAGAAAGGCTGAGCCCAAAAGACCCCGCGATGACGCG 1500
1835 GATTAAGCGAGCCAGCGGCTGCGCTCAGTGCAGGATCCATGAGCTGAGAGCGG 1894
1501 GATTAAGCGAGCCAGCGGCTGCGCTCAGTGCAGGATCCATGAGCTGAGAGCGG 1560
1895 GAGGAGCTCGGTGAGCTTGTGCGAGCAGGTACCAAAACAAATGTTCTGTCACGGGGC 1954
1561 GAGGAGCTCGGTGAGCTTGTGCGAGCAGGTACCAAAACAAATGTTCTGTCACGGGGC 1620
1955 ATGCTTCAATGCTGTTTCCCTGCAAGACATGAGAGAGTGAATGAATTTCAACAT 2014
1621 ATGCTTCAATGCTGTTTCCCTGCAAGACATGAGAGAGTGAATGAATTTCAACAT 1680
2015 TGCTTCAAGCAGGAGAGAGACTGTTCAAGTGTCTCCCGGCGTGTGAGATCTCAA 2074
1681 TGCTTCAAGCAGGAGAGAGACTGTTCAAGTGTCTCCCGGCGTGTGAGATCTCAA 1740
2075 CCGGTGCTCAAGAAAGAGAGCTTGTGCGCACTGTGCTTCAATCACTGCTGGGGCG 2134
1741 CCGGTGCTCAAGAAAGAGAGCTTGTGCGCACTGTGCTTCAATCACTGCTGGGGCG 1800
2135 GCTTCCGAGATGCTTGTGCTGCGCTGAGATCTGCTGAAAGCTGAGATGAGCTGTT 2194
1801 GCTTCCGAGATGCTTGTGCTGCGCTGAGATCTGCTGAAAGCTGAGATGAGCTGTT 1860
2195 TCTGAGCAATTA 2206
1861 TCTGAGCAATTA 1872

RESULT 14
US-09-953-351-12
; Sequence 12: Application US/09953351
; Publication No. US2003003643A1
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; APPLICANT: Melander, Christian
; APPLICANT: Liu, Hong-Xiang
; APPLICANT: Jin, Cheng He
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE CONSTRUCTION AND USE OF FUSION L
; FILE REFERENCE: A-70814/RFT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/953,351
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 60/232,960
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: adeno-associated virus 1
US-09-953-351-12

Query Match 39.7%; Score 1872; DB 12; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ATGCGGGCTTCTAGAGATCTGATCAAGTGGCCGAGCGCACTGAGCAGACCTTCCG 394
DB 1 ATGCGGGCTTCTAGAGATCTGATCAAGTGGCCGAGCGCACTGAGCAGACCTTCCG 60
QY 395 GGCATTTCTGACTGCTTGTGAGCTGGTGGCCGAGAGAAATGGAGTGGCCCGCGAT 454
DB 61 GGCATTTCTGACTGCTTGTGAGCTGGTGGCCGAGAGAAATGGAGTGGCCCGCGAT 120
QY 455 TCTGACATGATCTGAATCTGATTTAGACAGGACCCCTGACCGTGGCCGAGAACTTCCG 514
DB 121 TCTGACATGATCTGAATCTGATTTAGACAGGACCCCTGACCGTGGCCGAGAACTTCCG 180
QY 515 CGGCACTTCTGCTGCTCAATGAGCGCGCGTGAAGTAAAGCCCGGAGGCGCTTCTTTGTT 574
DB 181 CGGCACTTCTGCTGCTCAATGAGCGCGCGTGAAGTAAAGCCCGGAGGCGCTTCTTTGTT 240
QY 575 CAGTTGAGAGAGGCGGAGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 634
DB 241 CAGTTGAGAGAGGCGGAGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 300
QY 635 AAATCCATGCTGCTGAGCGCGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 694
DB 301 AAATCCATGCTGCTGAGCGCGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 360
QY 695 TACCGCGGATCGAGCGCGCTTCCCACTGTTCCGCGTGAACCAAGCGCTGATATGCG 754
DB 361 TACCGCGGATCGAGCGCGCTTCCCACTGTTCCGCGTGAACCAAGCGCTGATATGCG 420
QY 755 GCGGAGAGGAGGAGCAAGAGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 814
DB 421 GCGGAGAGGAGGAGCAAGAGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 480
QY 815 ACTGAGCGCGAGCTGAGTGGCGCTGAGCTTAACTAACTAACTAACTAACTAACTAACTAACTAACT 874
DB 481 ACTGAGCGCGAGCTGAGTGGCGCTGAGCTTAACTAACTAACTAACTAACTAACTAACTAACTAACT 540
QY 875 AACTGAGCGCGAGCGCAAGCGCTTGTGCGCGAGCACTGACCGCGTGAAGCGCGAG 934
DB 541 AACTGAGCGCGAGCGCAAGCGCTTGTGCGCGAGCACTGACCGCGTGAAGCGCGAG 600
QY 935 GAGCAGAAACAGAGAGTGTGAACCCCAATTCTGACGCGCTGTCACTCGGTCAGAAACC 994
DB 601 GAGCAGAAACAGAGAGTGTGAACCCCAATTCTGACGCGCTGTCACTCGGTCAGAAACC 660
QY 995 TCCGCGGCTACATGAGCTGTGTGGGT 1054

DB 661 TCCGCGGCTACATGAGCTGTGTGGGT 720
QY 1055 CAGTGTATCCAGAGAGACCAAGCGCTGTATCATCTCTTCAAGCGCGCTTCAACTGCGCG 1114
DB 721 CAGTGTATCCAGAGAGACCAAGCGCTGTATCATCTCTTCAAGCGCGCTTCAACTGCGCG 780
QY 1115 TCCAGATCAAGCGCGCTGTGACCAATGCGCGCAAGTCAATGCGCGTGAACCAATCGCG 1174
DB 781 TCCAGATCAAGCGCGCTGTGACCAATGCGCGCAAGTCAATGCGCGTGAACCAATCGCG 840
QY 1175 CCGGACTTCTGTTAGCGCGCGCTCGCGCGCGGACATTAAACCAACCGCATCTACCG 1234
DB 841 CCGGACTTCTGTTAGCGCGCGCTCGCGCGCGGACATTAAACCAACCGCATCTACCG 900
QY 1235 ATCTGAGCTGAACGCGCTACCAACCTGCTAGCGCGCTCTCTTCTCGGCTGCGCG 1294
DB 901 ATCTGAGCTGAACGCGCTACCAACCTGCTAGCGCGCTCTCTTCTCGGCTGCGCG 960
QY 1295 CAGAAAGGTTCCGGAAGCGCAACACCATCTGCTGTTGGGCGCGGCAACGCGGCAAG 1354
DB 961 CAGAAAGGTTCCGGAAGCGCAACACCATCTGCTGTTGGGCGCGGCAACGCGGCAAG 1020
QY 1355 ACCAATGCGGAGAGCATGCGCAAGCGCGCTTCTTCAAGCGCTGCAACTGCAAC 1414
DB 1021 ACCAATGCGGAGAGCATGCGCAAGCGCGCTTCTTCAAGCGCTGCAACTGCAAC 1080
QY 1415 AATGAGAACTTCTTCTTCAATGATTTGCTGTGAACAAATGATTTCTGTGTGTGTGTGTGTGTGTGT 1474
DB 1081 AATGAGAACTTCTTCTTCAATGATTTGCTGTGAACAAATGATTTCTGTGTGTGTGTGTGTGTGTGT 1140
QY 1475 AAGATGACGCGCAAGGCTGCTGAGTCCGCGCAAGGCACTTCTCGGCGGCAAGAGTGTGCG 1534
DB 1141 AAGATGACGCGCAAGGCTGCTGAGTCCGCGCAAGGCACTTCTCGGCGGCAAGAGTGTGCG 1200
QY 1535 GTGACCAAAAGTGCAAGTGTGCGCGCAAGTGAACCCCAACCGCGTGTGCTGCTCC 1594
DB 1201 GTGACCAAAAGTGCAAGTGTGCGCGCAAGTGAACCCCAACCGCGTGTGCTGCTCC 1260
QY 1595 AAGACCAACATGCGCGCGGATTTGAAGGGAAGCAACCTTTCAGAGCAAGAGCGCG 1654
DB 1261 AAGACCAACATGCGCGCGGATTTGAAGGGAAGCAACCTTTCAGAGCAAGAGCGCG 1320
QY 1655 TTGACAGACCGGAGTGTCAAAATTGAATCAACCGCGCTGTGAGAGTGAATTTGCAAG 1714
DB 1321 TTGACAGACCGGAGTGTCAAAATTGAATCAACCGCGCTGTGAGAGTGAATTTGCAAG 1380
QY 1715 GTGACAAAGGAGAGTCAAGAGTCTTCTCGTGGCGCGAGATCACTGACCGAGTG 1774
DB 1381 GTGACAAAGGAGAGTCAAGAGTCTTCTCGTGGCGCGAGATCACTGACCGAGTG 1440
QY 1775 GCGCATGATCTTCACTGATCAAGAAAGGTGAGCCCAAAAGACCGCGCGCGATGACGCG 1834
DB 1441 GCGCATGATCTTCACTGATCAAGAAAGGTGAGCCCAAAAGACCGCGCGCGATGACGCG 1500
QY 1835 GATTAAGGAGGCGCAAGCGGCGCTGCGCTGAGTGGCGGATCATCTGACAGCGCG 1894
DB 1501 GATTAAGGAGGCGCAAGCGGCGCTGCGCTGAGTGGCGGATCATCTGACAGCGCG 1560
QY 1895 GAAAGAGCTTCCGTTGAGCTTGTGCGACAGGTACCAAAACCAATGTTCTGTCAACGCGGCG 1954
DB 1561 GAAAGAGCTTCCGTTGAGCTTGTGCGACAGGTACCAAAACCAATGTTCTGTCAACGCGGCG 1620
QY 1955 ATGCTTCAATGCTGTTTCCCTGCAAGACATGGAAGAAATGAATGAATTTCAACAT 2014
DB 1621 ATGCTTCAATGCTGTTTCCCTGCAAGACATGGAAGAAATGAATGAATTTCAACAT 1680
QY 2015 TGCTTACGCAAGGAGAGAGTGTCAAGTGTCTTCCCGCGGCTGTCAAGTCTCAA 2074
DB 1681 TGCTTACGCAAGGAGAGAGTGTCAAGTGTCTTCCCGCGGCTGTCAAGTCTCAA 1740
QY 2075 CCGGTGTCAAGAAAGAGAGTATCGGAACTCTGTGCTCAATTCATTCATCTGCTGCGCGG 2134
DB 1741 CCGGTGTCAAGAAAGAGAGTATCGGAACTCTGTGCTCAATTCATTCATCTGCTGCGCGG 1800

Qy	2015	TGCTTCACGACCGGACGAGAGACTGTTCAAGAGTCTTCCCGGCGGTGACAGATCTCAA	2074
Db	1681	TGCTTCACGACGAGGACGAGAGACTGTTCAAGAGTCTTCCCGGCGGTGACAGATCTCAA	1740
Qy	2075	CCGGTCGTCAAGAAAGAGACGTATCGGAACTCTGTGCCATTCATCTGCTGGGCGG	2134
Db	1741	CCGGTCGTCAAGAAAGAGACGTATCGGAACTCTGTGCCATTCATCTGCTGGGCGG	1800
Qy	2135	GCTCCCGAGATGCTTGTCTGGGCTGCGATCTGTCACAGTGAACCTGAGTGAATGACTGT	2194
Db	1801	GCTCCCGAGATGCTTGTCTGGGCTGCGATCTGTCACAGTGAACCTGAGTGAATGACTGT	1860
Qy	2195	TCTGAGCAATTA	2206
Db	1861	TCTGAGCAATTA	1872

Search completed: July 20, 2003, 20:11:27
 Job time : 1032.56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 12440.8 Seconds

(without alignments)
11036.790 Million cell updates/sec

Title: US-09-807-802A-1

Perfect score: 4718

Sequence: 1 ttgcacacccctctctgcg.....cgacagagagtggtggcaaa 4718

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4718	100.0	4718	14	AF063497
2	4258	90.3	4663	14	AF028704
3	3211.6	68.1	4726	14	AF028704
4	3205	67.9	4722	14	AF028705
5	3107.8	65.9	4679	6	AX282480
6	3107.8	65.9	4679	14	AF043303
7	3107.8	65.9	4679	6	AX205072
8	3082.4	65.3	4680	6	AR028767
9	3082.4	65.3	4680	6	162303
10	3055.8	64.8	4675	6	AX135805
11	3055.8	64.8	4675	6	AX286292
12	3055.8	64.8	4675	14	AA2CG
13	2993.4	63.4	7557	6	AX205073
14	2878.4	61.0	7327	12	AF369963
15	2598.6	55.1	4767	14	AVU89790
16	1691.2	35.8	4072	6	AX205074
17	1472.2	31.2	2116	14	AA2LEPT
18	1425.8	30.2	4910	6	AR027100
19	1425.8	30.2	4910	6	AR036062
20	1425.8	30.2	4910	6	AR160052
21	1425.8	30.2	4910	6	AR209764
22	1354.2	28.7	4642	14	AF085716
23	1354.2	28.7	4652	6	AX256321
24	1352.6	28.7	4404	14	AAV18065
25	1008	21.4	9258	12	AF383623
26	877.2	18.6	4101	14	AF416726
27	861.6	18.3	5132	14	BDU22967
28	860	18.2	5106	14	GPU25749
29	852	18.1	4550	14	MDPNSVP
30	795.8	16.9	3434	14	GPU34761
31	773.4	15.6	2307	6	AX256327
32	729.8	15.5	2284	6	AX256328
33	729.8	15.5	2264	6	AX256329
34	690.6	14.6	1690	6	AX256335
35	690.6	14.6	1870	6	AX256331
36	616	13.1	2487	14	MDPVPS
37	476	10.1	1115	6	AX256333
38	475	10.1	1292	6	AX256330
39	433.8	9.2	606	6	AX344105
40	399	8.5	653	14	XXAAV1
41	262.8	5.6	579	14	HMAAV01
42	229.2	4.9	539	14	AF166104
43	229.2	4.9	539	14	AF166105
44	229.2	4.9	539	14	AF166106
45	227.6	4.8	539	14	AF166102

ALIGNMENTS

```
RESULT 1
LOCUS AF063497 4718 bp DNA linear VRL 27-APR-1999
DEFINITION Adeno-associated virus 1, complete genome.
ACCESSION AF063497
VERSION AF063497.1 GI:4689096
KEYWORDS
ORGANISM adeno-associated virus 1.
REFERENCE 1 (bases 1 to 4718)
AUTHORS Xiao,W., Chirmule,N., Berta,S.C., McCullough,B., Gao,G. and
Wilsen,J.M.
TITLE Gene therapy vectors based on adeno-associated virus type 1
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)
```

MEDLINE 99214338
 PUBMED 10196295
 REFERENCE 2 (bases 1 to 4718)
 AUTHORS Xiao, W. and Wilson, J.M.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA

FEATURES
 SOURCE location/Qualifiers

CDS

1. 4718
 /organism="adenovirus-1"
 /db_xref="taxon:85106"
 /note="AAV1"
 335..2206
 /codon_start=1
 /product="nonstructural protein"
 /protein_id="A027758.1"
 /db_xref="gi:4689098"
 /translation="MPGFYELIVKPSDLDLHPLGSDSPVSWAEKEMWLPDSDMD
 LNLLEQAPLVAKLQDPLVQMRVSKAPALPFOPEKESYFHLILVETGVGS
 MVLGRPLSLQIDKLVTIYRIGLEPLPMPVATKTRNGAGGKLVDECVIPYLK
 TQPELQWMTMEYISACNLARKEVLVQHLTVSOTONKEMINPNSDAVIRS
 KTSARMEVLVDRGITSKQNIQEDQASYFNASNSRSQIKALDNAGITMAL
 TKSAADYLVGAPPADIKTRNIYRILELNGEPYASVFLGMQKPGKXNTIMLG
 PATGKTNIAEAIHAAVFGCVMTNENPFPCVDKMTWMEGKMTAKVESA
 ILGSKRVDOCKSSAIDPTPIVTSNTMCAVIDGNTTFHQDPLDPMKFEEL
 TRLEHDEGKVTQKEKFEFRMAODHTEVHEPFAKPKAPDADKSEPKA
 CPSVADPSTDAEDAPVDPRADYONKCSRHAGMLPCKTCRMONNPTCTHGT
 RDGSECFPGVSESQPVAKRTYRKLCAIHLGLAPLACACDLVNDLDDCVSEQ"
 2223..4433
 /codon_start=1
 /product="capsid protein"
 /protein_id="A027757.1"
 /db_xref="gi:4689097"
 /translation="MAADGYLPDLLEMLSEGRFEMWDLKPKAPKANOQKODGK
 LVLGKYLGPNGLDKGEVYNAADALIEDKAYDQKAGDNPYRINADAEFPE
 RLQEDTSFGNIGRAVLEPQAKRVLLELVEEGATVAKRPEVQSPQEDSSGIG
 KTGQOPAKRLNFGQDSEVPDPOPLGEPATVAVGPTTASGGAGPADNBEA
 DGVNASGNHMDSTWLDGRVITLSTFWALPTVNNHLYKQISASTGASNDNYFGY
 STPWGVPFNRHCHFSRDMOQLINNNWGRPKRLNKLKFNLYLTINNGOAVRTTA
 NNLTSTVOVDSERYQPYVIGSAHGLPPRPADVPIPIQYVLTINNGOAVRTTA
 FYCLTFPSQMLRTGNNTFSYFEEVHFSYAHOSLDRLANPLIDQYLYNTRQ
 NQSSAONKDLFSGSPAGMSVOPKMLPGCYQRQSVSTKTDNNNSNTYNAK
 YNLNGRESILNPGTAMASHKDEDEKFFPMGSMVTFKGSASASTVLDNWIIDEEI
 KATNPATERGTVAVNFOSSTDPATDEHAMGLPDMVQDDVLYOGEIMAKTTH
 TDGHPHPSKLMGSGELKNPPOILLIKNPVNPANPAESATKRPASFTIYVSTGVSE
 IEMLEQKENSRRNPBYVYTSYAKSANVDTVNNGLYTERPIGRYILRPL"

CDS

BASE COUNT 1121 a 1393 c 1273 g 931 t
 ORIGIN

Query Match 100.0%; Score 4718; DB 14; Length 4718;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTGCCACTCCCTCTCTGCGCGCTGCTCGCTCGGTGGGCGCTTGGACCAAGGCTCCG 60
 1 TTGCCACTCCCTCTCTGCGCGCTGCTCGCTCGGTGGGCGCTTGGACCAAGGCTCCG 60
 1 AGAGCGGAGAGCTCTGCTGCGCGCGCCCAACGACGAGCGGCGGAGAGGAGATG 120
 61 AGAGCGGAGAGCTCTGCTGCGCGCGCCCAACGACGAGCGGCGGAGAGGAGATG 120
 61 AGAGCGGAGAGCTCTGCTGCGCGCGCCCAACGACGAGCGGCGGAGAGGAGATG 120
 121 GCGCAATCCATCTAGGGGTATAGCGAAGCGCTCCCAACGCTGCGCGCTGAGCGCTGA 180
 121 GCGCAATCCATCTAGGGGTATAGCGAAGCGCTCCCAACGCTGCGCGCTGAGCGCTGA 180
 121 GCGCAATCCATCTAGGGGTATAGCGAAGCGCTCCCAACGCTGCGCGCTGAGCGCTGA 180
 181 CGTAATATTACGTCATAGGGAGTGTCTCTGATATAGCTGTCACTGAGTGTCTTTGGAC 240
 181 CGTAATATTACGTCATAGGGAGTGTCTCTGATATAGCTGTCACTGAGTGTCTTTGGAC 240
 241 ATTTTGGACACCAAGTGGCATTTAAGGTATATATGCGCGAGTGAAGGACAGATCTC 300
 241 ATTTTGGACACCAAGTGGCATTTAAGGTATATATGCGCGAGTGAAGGACAGATCTC 300

301 CATTTGACCGGAAATTTGACGAGCAGCAGCATCCGCGCTTCTACAGATCGTAT 360
 301 CATTTGACCGGAAATTTGACGAGCAGCAGCATCCGCGCTTCTACAGATCGTAT 360
 361 CAAGTCCGAGGAGCTTGGACGAGCAGCTTCCGGGCTTTCTGACTCGTTTGTAGCTG 420
 361 CAAGTCCGAGGAGCTTGGACGAGCAGCTTCCGGGCTTTCTGACTCGTTTGTAGCTG 420
 421 GGTGGCCGAGAGAAATGGAGAGTGGCCCGCGGATTCGATCATGATCTGATTTGA 480
 421 GGTGGCCGAGAGAAATGGAGAGTGGCCCGCGGATTCGATCATGATCTGATTTGA 480
 481 GCAAGCAGCCCTGACCGTGGCCGAGAGAGTGAAGCGGACCTTCTGATTCGATTCGAT 540
 481 GCAAGCAGCCCTGACCGTGGCCGAGAGAGTGAAGCGGACCTTCTGATTCGATTCGAT 540
 541 CGTGAATAGGCGCCCGGAGCGCTTCTTCTTGTTCAGTTGAGAGAGGCGAGTCTTACT 600
 541 CGTGAATAGGCGCCCGGAGCGCTTCTTCTTGTTCAGTTGAGAGAGGCGAGTCTTACT 600
 601 CCACTTCATATTCTGCTGGTGAAGCAAGGCGGCTCAATCCATGCTGCTGGCGCTTCT 660
 601 CCACTTCATATTCTGCTGGTGAAGCAAGGCGGCTCAATCCATGCTGCTGGCGCTTCT 660
 661 GAGTCAGATTAGGAGCAAGCTGTGTGAGCAATCTACCGGGAGATGAGCCGACCTTCC 720
 661 GAGTCAGATTAGGAGCAAGCTGTGTGAGCAATCTACCGGGAGATGAGCCGACCTTCC 720
 661 GAGTCAGATTAGGAGCAAGCTGTGTGAGCAATCTACCGGGAGATGAGCCGACCTTCC 720
 721 CAATGCTTGGCGGTGACCAAGCGCTAATGGCGCGGAGGGGGGAAAGGTTGTGGA 780
 721 CAATGCTTGGCGGTGACCAAGCGCTAATGGCGCGGAGGGGGGAAAGGTTGTGGA 780
 781 CGAGTGTACATCTCCCACTACCTCTGCGCCAGAGACTCAAGCGAGCTGAGTGGCGGT 840
 781 CGAGTGTACATCTCCCACTACCTCTGCGCCAGAGACTCAAGCGAGCTGAGTGGCGGT 840
 841 GACTTAATGAGAGAGATATATTAAGCGCTGTTGAACCTGGCGGAGCGCAAGCGCTGT 900
 841 GACTTAATGAGAGAGATATATTAAGCGCTGTTGAACCTGGCGGAGCGCAAGCGCTGT 900
 901 GGGCGACACCTGACCTGACCTGACCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAG 960
 901 GGGCGACACCTGACCTGACCTGACCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAG 960
 961 CAATTTGAGCGCGCTCTCATCCGCTCAAAAACCTCCGCGCTGATCATGAGCTGTGCG 1020
 961 CAATTTGAGCGCGCTCTCATCCGCTCAAAAACCTCCGCGCTGATCATGAGCTGTGCG 1020
 1021 GTGGCTGTGAGACCGGGGATCACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 1021 GTGGCTGTGAGACCGGGGATCACTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 1081 GTATATCTCTTCAAGCGCGCTTCCAACTGGCGGTCCCAATGAGAGAGAGAGAGAGAG 1140
 1081 GTATATCTCTTCAAGCGCGCTTCCAACTGGCGGTCCCAATGAGAGAGAGAGAGAGAG 1140
 1141 TGCCGGAAGATCATGCGCGCTTCCAACTGGCGGTCCCAATGAGAGAGAGAGAGAGAG 1200
 1141 TGCCGGAAGATCATGCGCGCTTCCAACTGGCGGTCCCAATGAGAGAGAGAGAGAGAG 1200
 1201 GCGCGGAGATTAACCAACCGCATCTACCGCATCTGAGAGTGAAGCGGCTTCAAGAC 1260
 1201 GCGCGGAGATTAACCAACCGCATCTACCGCATCTGAGAGTGAAGCGGCTTCAAGAC 1260
 1261 TGCTTAAGCGCGCTCTTCTGCGGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 1261 TGCTTAAGCGCGCTCTTCTGCGGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 1321 CATCTGCTGTTTGGCGCGGACCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 1321 CATCTGCTGTTTGGCGCGGACCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 1381 CGCGGTGCTTCTTCAAGCGGTGCTGCACTGAGACCAATGAGAGAGAGAGAGAGAG 1440

Db	1381	CGCGTCCCTTCAACGGCTGCGCACTGGACCAATGAGAACTTCCCTTCAATGATG	1440
Qy	1441	CGTCGAACAATGGTATCTGTGGGAGAGGCGAATATGACGGCCCAAGTCTGGAGTC	1500
Db	1441	CGTCGAACAATGGTATCTGTGGGAGAGGCGAATATGACGGCCCAAGTCTGGAGTC	1500
Qy	1501	CGCGAAGGCAATCTCGGCGGAGAGAGTGGCGTGAACCAAAAGTGAAGTCTGCGC	1560
Db	1501	CGCGAAGGCAATCTCGGCGGAGAGAGTGGCGTGAACCAAAAGTGAAGTCTGCGC	1560
Qy	1561	CGAATGACCCCAACCCCGGTGATGTCACCTCCACCAACCAATGTCGCGGTGATGA	1620
Db	1561	CGAATGACCCCAACCCCGGTGATGTCACCTCCACCAACCAATGTCGCGGTGATGA	1620
Qy	1621	CGGGAACAGACCACTTCCAGACCAAGACCGGTGACAGACCGGATGTTCAATTGA	1680
Db	1621	CGGGAACAGACCACTTCCAGACCAAGACCGGTGACAGACCGGATGTTCAATTGA	1680
Qy	1681	ACTCACCCGCGTCTGGAGCATGACTTGGCAAGGTGACAAAGAGGAAGTCAAGATT	1740
Db	1681	ACTCACCCGCGTCTGGAGCATGACTTGGCAAGGTGACAAAGAGGAAGTCAAGATT	1740
Qy	1741	CTTCCGCTGCGGAGATGACGTCAGAGGTGGCGCATGATGTTCTACGTCAAGAAAGG	1800
Db	1741	CTTCCGCTGCGGAGATGACGTCAGAGGTGGCGCATGATGTTCTACGTCAAGAAAGG	1800
Qy	1801	TGGAGCCCAACAAAGACCCCGGATGACGCGGATTAAGAGAGCCCAAGCGGCGCTG	1860
Db	1801	TGGAGCCCAACAAAGACCCCGGATGACGCGGATTAAGAGAGCCCAAGCGGCGCTG	1860
Qy	1861	CCCTCAGTGGCGGATCATGACGTGACAGCGGAGAGAGTCCGCTGAGCTTTGCGCA	1920
Db	1861	CCCTCAGTGGCGGATCATGACGTGACAGCGGAGAGAGTCCGCTGAGCTTTGCGCA	1920
Qy	1921	CAGGTACCAAAACAAATGTTCTGTCACGCGGCGATGTTCAATGCTTTCCCTGCA	1980
Db	1921	CAGGTACCAAAACAAATGTTCTGTCACGCGGCGATGTTCAATGCTTTCCCTGCA	1980
Qy	1981	GACATGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2040
Db	1981	GACATGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2040
Qy	2041	TTGAGATGCTTCCCGCGGCTGACAGATCTCAACCGGTGTCAGAAAGAGAGTATCG	2100
Db	2041	TTGAGATGCTTCCCGCGGCTGACAGATCTCAACCGGTGTCAGAAAGAGAGTATCG	2100
Qy	2101	GAAACTCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2160
Db	2101	GAAACTCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2160
Qy	2161	CGATCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2220
Db	2161	CGATCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2220
Qy	2221	GATGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2280
Db	2221	GATGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2280
Qy	2281	GCGAGTGTGGAATTGAAACCTGGAACCCCGAAGCCCAAGCCCAAGCAAAAGCAGG	2340
Db	2281	GCGAGTGTGGAATTGAAACCTGGAACCCCGAAGCCCAAGCCCAAGCAAAAGCAGG	2340
Qy	2341	ACGACGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2400
Db	2341	ACGACGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2400
Qy	2401	ACAAAGGAGAGCCGCTGACAGCGGCGAGCGAGCGGCGCTGAGCAAGAGGCTTACG	2460
Db	2401	ACAAAGGAGAGCCGCTGACAGCGGCGAGCGAGCGGCGCTGAGCAAGAGGCTTACG	2460
Qy	2461	ACGAGAGCTTCAAGCGGCTGACCAATCTGTCCTGCGGTATTAACAGCGCGAGT	2520
Db	2461	ACGAGAGCTTCAAGCGGCTGACCAATCTGTCCTGCGGTATTAACAGCGCGAGT	2520
Db	2461	ACGAGAGCTTCAAGCGGCTGACCAATCTGTCCTGCGGTATTAACAGCGCGAGT	2520
Qy	2521	TTGAGAGGCTGCAAGAAATAGTCTTTTGGGGCAACCTCGGCGAGAGTCTTCC	2580
Db	2521	TTGAGAGGCTGCAAGAAATAGTCTTTTGGGGCAACCTCGGCGAGAGTCTTCC	2580
Qy	2581	AGGCAAGAGCGGCTTCTGAACTCTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTG	2640
Db	2581	AGGCAAGAGCGGCTTCTGAACTCTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTG	2640
Qy	2641	CTGGAAGAAAGCTCGGCTGAGACAGTCCGCAAGAGCCAGCTCTCTCGGCGATCG	2700
Db	2641	CTGGAAGAAAGCTCGGCTGAGACAGTCCGCAAGAGCCAGCTCTCTCGGCGATCG	2700
Qy	2701	GCAAGAGGCGCAGAGCCCGCTTAAAGAGAGCTCAATTTGATGACAGTGGAGCTGAG	2760
Db	2701	GCAAGAGGCGCAGAGCCCGCTTAAAGAGAGCTCAATTTGATGACAGTGGAGCTGAG	2760
Qy	2761	AGTCAAGTCCCGATCCACAACTCTGAGAACTCTCAGCAACCCCGCTGCTGAGGAC	2820
Db	2761	AGTCAAGTCCCGATCCACAACTCTGAGAACTCTCAGCAACCCCGCTGCTGAGGAC	2820
Qy	2821	CTACTCAATGCTTCAAGCGGCTGAGCGGCGCACTTGGCAGCAATTAAGAGGCGCGACG	2880
Db	2821	CTACTCAATGCTTCAAGCGGCTGAGCGGCGCACTTGGCAGCAATTAAGAGGCGCGACG	2880
Qy	2881	GAGTGGGTATGCTCAGAAATTTGGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG	2940
Db	2881	GAGTGGGTATGCTCAGAAATTTGGCAATTTGCAATTTGCAATTTGCAATTTGCAATTTG	2940
Qy	2941	TCAACACAGACACCCGACCTGCGGCTTGGCCACTTCAATTAACCACTTCAAGCAAA	3000
Db	2941	TCAACACAGACACCCGACCTGCGGCTTGGCCACTTCAATTAACCACTTCAAGCAAA	3000
Qy	3001	TTCTCAGTCTTCAAGCGGCGGCGAGCAAGCACTTCTGCTGACGACCCCTT	3060
Db	3001	TTCTCAGTCTTCAAGCGGCGGCGAGCAAGCACTTCTGCTGACGACCCCTT	3060
Qy	3061	TCTCAGTCTTCAAGCGGCGGCGAGCAAGCACTTCTGCTGACGACCCCTT	3120
Db	3061	TCTCAGTCTTCAAGCGGCGGCGAGCAAGCACTTCTGCTGACGACCCCTT	3120
Qy	3121	TCATCAACAACATTTGGGATTCGCGCCAGAGACTCAACTTCAACATCC	3180
Db	3121	TCATCAACAACATTTGGGATTCGCGCCAGAGACTCAACTTCAACATCC	3180
Qy	3181	AACTCAAGAGGTACAGACGATGATGATGATGATGATGATGATGATGATGATGATGATG	3240
Db	3181	AACTCAAGAGGTACAGACGATGATGATGATGATGATGATGATGATGATGATGATGATG	3240
Qy	3241	CGGTTCAAGTCTTCTGGAATCGGAGTACAGCTTCCGTCGCTGCTGCTGCTGCTGCTG	3300
Db	3241	CGGTTCAAGTCTTCTGGAATCGGAGTACAGCTTCCGTCGCTGCTGCTGCTGCTGCTG	3300
Qy	3301	AGGCTGCTCCTCTCGTTCGCGGAGAGTGTTCATGATTCGCAATAAGCTTACCTGA	3360
Db	3301	AGGCTGCTCCTCTCGTTCGCGGAGAGTGTTCATGATTCGCAATAAGCTTACCTGA	3360
Qy	3361	CGCTCAACAAATGAGCAAGCCGCTGAGAGTTCATCTTTTATCTGCTGGAATATTC	3420
Db	3361	CGCTCAACAAATGAGCAAGCCGCTGAGAGTTCATCTTTTATCTGCTGGAATATTC	3420
Qy	3421	CTTCTCAAGTCTGAGAAAGCGGCAACCTTACCTTCAAGTCTTCAAGTCTTCAAGTCTT	3480
Db	3421	CTTCTCAAGTCTGAGAAAGCGGCAACCTTACCTTCAAGTCTTCAAGTCTTCAAGTCTT	3480
Qy	3481	CTTTCACAGCACTACGCGACAGCGAGCTGAGCGGCTGATGATCTCTCATCG	3540
Db	3481	CTTTCACAGCACTACGCGACAGCGAGCTGAGCGGCTGATGATCTCTCATCG	3540
Qy	3541	ACCAATACCTGTATTAATCTTGAACAGAACTTCAATTAATGATGATGATGATGATGATG	3600
Db	3541	ACCAATACCTGTATTAATCTTGAACAGAACTTCAATTAATGATGATGATGATGATGATG	3600

QY 3601 ACTTGCTGTTAGCGCGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAAACTGGCTAC 3660
 DB 3601 ACTTGCTGTTAGCGCGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAAACTGGCTAC 3660
 QY 3661 CTGACCCCTGTTATCGGAGCAGCGCGTCTTAAACAAAAACAGACAAACAGACGA 3720
 DB 3661 CTGACCCCTGTTATCGGAGCAGCGCGTCTTAAACAAAAACAGACAAACAGACGA 3720
 QY 3721 ATTTTACCTGGAAGTGTCTTCAAAATTAACCTCAATGGCGTGAATCCATCATCAAC 3780
 DB 3721 ATTTTACCTGGAAGTGTCTTCAAAATTAACCTCAATGGCGTGAATCCATCATCAAC 3780
 QY 3781 CTGGACCTGTAATGAGCTCACAACAAAGCAGAAAGCAAGTCTTTCCATGAGCGGTG 3840
 DB 3781 CTGGACCTGTAATGAGCTCACAACAAAGCAGAAAGCAAGTCTTTCCATGAGCGGTG 3840
 QY 3841 TCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAAACTGCAATGGACAATGCTATGA 3900
 DB 3841 TCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAAACTGCAATGGACAATGCTATGA 3900
 QY 3901 TTACAGACGAAAGAGAAATTAAGCACTTAACCTGTGGCCACCGAAAGATTGGACCG 3960
 DB 3901 TTACAGACGAAAGAGAAATTAAGCACTTAACCTGTGGCCACCGAAAGATTGGACCG 3960
 QY 3961 TGGGACCAATTTCCAGAGCAGACAGACAGACCGCTGGGAGCGGAGATGTGATGCTATG 4020
 DB 3961 TGGGACCAATTTCCAGAGCAGACAGACAGACCGCTGGGAGCGGAGATGTGATGCTATG 4020
 QY 4021 GAGCATTAACCTGGAAGTGTGGCAAGATAGAGAGTGTACCTGCAAGGCTCCATTTGGG 4080
 DB 4021 GAGCATTAACCTGGAAGTGTGGCAAGATAGAGAGTGTACCTGCAAGGCTCCATTTGGG 4080
 QY 4081 CCAAAATTTCTCACAACAGATGGAACCTTCAACCGCTCTCTTATGGCGCGCTTTGGAC 4140
 DB 4081 CCAAAATTTCTCACAACAGATGGAACCTTCAACCGCTCTCTTATGGCGCGCTTTGGAC 4140
 QY 4141 TCAAGAACCCGCTCTCAGATCTCATCAAAACAGCGCTGTCTGCGAATCCTCCGG 4200
 DB 4141 TCAAGAACCCGCTCTCAGATCTCATCAAAACAGCGCTGTCTGCGAATCCTCCGG 4200
 QY 4201 CGGAGTTTCCAGCTACAAAGTTTGTTCATTCATCAACCAATCTCCAAGACAGTGA 4260
 DB 4201 CGGAGTTTCCAGCTACAAAGTTTGTTCATTCATCAACCAATCTCCAAGACAGTGA 4260
 QY 4261 GTGAGAAATTTGATGGAGCTGCAAGAAACAGCAACGCTGGAATCCGAAAGTGC 4320
 DB 4261 GTGAGAAATTTGATGGAGCTGCAAGAAACAGCAACGCTGGAATCCGAAAGTGC 4320
 QY 4321 AGTACACATCCCAATTTATGCAAAATCTGCAACGTTGATTTTATCTGTGACAAACATGAC 4380
 DB 4321 AGTACACATCCCAATTTATGCAAAATCTGCAACGTTGATTTTATCTGTGACAAACATGAC 4380
 QY 4381 TTTATCTGAGCCCTCGCCCAATGGACCGCTTACCTTACCGTCTGTAATTAACGTG 4440
 DB 4381 TTTATCTGAGCCCTCGCCCAATGGACCGCTTACCTTACCGTCTGTAATTAACGTG 4440
 QY 4441 TTTATCTGAGCCCTCGCCCAATGGACCGCTTACCTTACCGTCTGTAATTAACGTG 4500
 DB 4441 TTTATCTGAGCCCTCGCCCAATGGACCGCTTACCTTACCGTCTGTAATTAACGTG 4500
 QY 4501 TATGCTGTTACGATGTTATGCTTACACATTAACCTGCTGTGCTGCGCATTAAG 4560
 DB 4501 TATGCTGTTACGATGTTATGCTTACACATTAACCTGCTGTGCTGCGCATTAAG 4560
 QY 4561 ACTTACGTCATCGGTTACCCCTAGTATGAGAGTTGCCCATCTCTGCGGCTGCG 4620
 DB 4561 ACTTACGTCATCGGTTACCCCTAGTATGAGAGTTGCCCATCTCTGCGGCTGCG 4620
 QY 4621 TCGCTCGTGGGCGCTGCGGACAAAGGTCCGACAGCGGACGACTCTGCTGCGCGCC 4680
 DB 4621 TCGCTCGTGGGCGCTGCGGACAAAGGTCCGACAGCGGACGACTCTGCTGCGCGCC 4680

QY 4681 CCACCGAGCGAGCGCGCAGAGAGGAGTGGCAA 4718
 DB 4681 CCACCGAGCGAGCGCGCAGAGAGGAGTGGCAA 4718
 RESULT 2
 AF028704
 LOCUS
 DEFINITION
 Adeno-associated virus 6, complete genome.
 ACCESSION
 AF028704
 VERSION
 AF028704.1 GI:2766605
 KEYWORDS
 SOURCE
 ORGANISM
 adeno-associated virus 6.
 adeno-associated virus 6
 Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 REFERENCE
 1 (bases 1 to 4683)
 Rutledge, E.A., Halbert, C.L. and Russell, D.W.
 Infectious clones and vectors derived from adeno-associated virus
 (AAV) serotypes other than AAV type 2
 J. Virol. 72 (1), 309-319 (1998)
 JOURNAL
 MEDLINE
 PUBMED
 9420229
 REFERENCE
 2 (bases 1 to 4683)
 Rutledge, E.A. and Russell, D.W.
 Direct Submission
 Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box
 357720, Seattle, WA 98195, USA
 FEATURES
 source
 1..4683
 /organism="adeno-associated virus 6"
 /db_xref="taxon:68558"
 /note="AAV6"
 320..2191
 /codon_start=1
 /evidence=not_experimental
 /product="nonstructural protein Rep78"
 /protein_id="AAB95449.1"
 /db_xref="GI:2766606"
 /translation="MPGYEIVIKVPSDIDELHFGISDSFVNMVAKEWELPPSDMD
 LNIIEBAPITVAEKORDFLVQMRVSKAPALFPQPEKGSFYHILHIVETGVKS
 MVIGRLSQIRDLVQITIRGIEPLNFAATKTRNGAGGKRVDECIITNYLLPK
 TOPELQAMTINMEVYISACLINAEERLVAHDTIVSOTQEDKNEINSDAPVIRS
 KTSARVMEVLGMLVDRGITSEKQMIQEDQASYISFVAASNSOIKALDNGKIMAL
 KTSAPRYLVGAPADIKTNRIYRIELGVDPAVAYSGFLGAOKRPEKRTIWLFG
 PATTGKTNTAIAHAPVPGVGCNMENEPDVCYDKVIMWESKMTAKVESAKA
 IIGGSVVRVDOCKSSAOIDPPIVTVSTNNCAVLDGKSTFEEHOPIODMEKEL
 TRLEHDFGKVTQEVKEFFRMAODHVTVAHEFYRKGAANKRPAPDDADSEPGA
 CBSVADPSTSDAGAVPADRYONKCSHAGLQMLFCKTCERNQNFNICFTGT
 RDCSECFPGVSESPVVRKRYRKLCAIHHLGRAPELIASACDILVNDLDCVSEQ"
 2208..4418
 /codon_start=1
 /evidence=not_experimental
 /product="capsid protein VP1"
 /protein_id="AAB95450.1"
 /db_xref="GI:2766607"
 /translation="MAADGYLPMLENDLSEGIKRWMDLPGAPKPKANOKODDGRG
 LVLPYKYLGVNGLDKGEPVNAADAADLAHEHKAAYQOLKAGDNPYLRYNHADABEG
 RLQDEPSRCGNGRAVPOAKRVLRBPGLVGEKATQALPKKRPVQESPOEPPSSSGIG
 KTGQOPAKRLNFGQIGSESVDDPOLPEPPATPAVPTTMASSGGA PMADNNGA
 DVGNGNSGWHDCDSTWLDGRVITSTRVALPTNNHLYKQISSASTGSDNHNFGY
 STPMWGFDFNRPFCHEPSRDMORLNNNGGFPKNNHFLFNIOVEAVTNGVTTIA
 NNLSTVOVFSSEVLPVYLSAHQGCLEPPADVFMIPOGYLTLLNGSAVGRS
 FYCLEYFSPQMRITGNPFSTYFEDPVSAYHOSQDLPLNPLIIDYLYLVNRSO
 NQSGSKNDLPLSRSPACQVOPKMWPGSCVIRQOARSKTKTNNNSNFMVTGASK
 YNLNGRESITINETSAMASHKDDKDFPDSGVMIKESKASNTLALDVMITDEEIK
 KATNPVATERFETVAVNLQSSSTDPATGVHGWALPGWVQDADVYLGSPIMAKIPH
 TGHFHPSPIMGFGFLKHPPOILINTEVPANPAPAEFSAITKFASTITQYSTGOVSE
 IEWELOKEMSKWNPDEVQYTSNVAKSANDFTVNDNGLYTEBRPIGRTYLTBPL"
 BASE COUNT 1114 a 1363 c 1277 g 929 t
 ORIGIN
 Query Match 90.3%; Score 4258; DB 14; Length 4683;
 Best Local Similarity 94.7%; Pred. No. 0;

Matches 4474; Conservative 0; Mismatches 205; Indels 43; Gaps 5;

QY 1 TTGCCCCTCTCTTCTGCGCGCTCGCTCGCTGCGGAGCTGCGGAGCCAAAGGTCCGC 60
DB 1 TTGGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTAGAGCGCGGACCAAAAGGTCCGC 60

QY 61 AGACGCGAGAGCTGCTGCTGCGCGCGCCCAACGAGCGAGGAGCGGAGGAGGAGT 120
DB 61 CGAGCGCCGCGGCTTTGCGCGCGCGCTCACTGAGCGAGCGAGCGGAGGAGGAGT 120

QY 121 GGCACATCTCACTAGAGGAGTATCGGAAAGCGCTCCCAAGCTCCGCTGACGCTGA 180
DB 121 GCAACATCTCACTAGAGGAGTATCGGAAAGCGCTCCCAAGCTCCGCTGACGCTGA 180

QY 181 CGTAAATTAAGTCACTAGG---GAGTGTCTCTGATTAAGCTGTCACTGAGTCTTTGC 237
DB 181 CGTAAATTAAGTCACTAGG---GAGTGTCTCTGATTAAGCTGTCACTGAGTCTTTGC 237

QY 164 CGTAAATTAAGTCACTAGGAGTATCGGAAAGCGCTCCCAAGCTCCGCTGACGCTGA 222
DB 164 CGTAAATTAAGTCACTAGGAGTATCGGAAAGCGCTCCCAAGCTCCGCTGACGCTGA 222

QY 238 GACATTTTGGAGACCAAGTGGCCATTAGGGTATATGAGCGAGTGAAGAGGAGGAGT 297
DB 238 GACATTTTGGAGACCAAGTGGCCATTAGGGTATATGAGCGAGTGAAGAGGAGGAGT 297

QY 223 GACATTTTGGAGACCAAGTGGTCAAGCTGAGTATTTAAGCGGAGTGAAGAGGAGT 282
DB 223 GACATTTTGGAGACCAAGTGGTCAAGCTGAGTATTTAAGCGGAGTGAAGAGGAGT 282

QY 298 CTCATTTTGAACCGGAAATTTGAACGAGAGCAGCAGCAGTCCGCGCTTCTAGAGAGT 357
DB 298 CTCATTTTGAACCGGAAATTTGAACGAGAGCAGCAGCAGTCCGCGCTTCTAGAGAGT 357

QY 283 CTCATTTTGAAGCGGAGGTTTGAAGCGGAGCAGTCCGCGCTTCTAGAGAGT 342
DB 283 CTCATTTTGAAGCGGAGGTTTGAAGCGGAGCAGTCCGCGCTTCTAGAGAGT 342

QY 358 GATCAAGTGGCGAGCGAGCTTGAAGAGCAGTCCGCGGATTTCTGAGCTGCTTGTGAG 417
DB 358 GATCAAGTGGCGAGCGAGCTTGAAGAGCAGTCCGCGGATTTCTGAGCTGCTTGTGAG 417

QY 343 GATTAAAGTCCCGGACCTTGAAGAGCAGTCCGCGGATTTCTGAGCAGTCTTGTGAA 402
DB 343 GATTAAAGTCCCGGACCTTGAAGAGCAGTCCGCGGATTTCTGAGCAGTCTTGTGAA 402

QY 418 CTGGGTTGGCGAGAGAGTGGAGCTGCGCGGATTTGAGCACTGAGTCTGAT 477
DB 418 CTGGGTTGGCGAGAGAGTGGAGCTGCGCGGATTTGAGCACTGAGTCTGAT 477

QY 403 CTGGGTTGGCGAGAGAGTGGAGTGGCGGAGTCTGAGCAGTCTGAGTCTGAT 462
DB 403 CTGGGTTGGCGAGAGAGTGGAGTGGCGGAGTCTGAGCAGTCTGAGTCTGAT 462

QY 478 TGAGCAGCAGCCTCTGACCGTGGCGGAGAGCTGACGCGGACTTCTGATCTGAT 537
DB 478 TGAGCAGCAGCCTCTGACCGTGGCGGAGAGCTGACGCGGACTTCTGATCTGAT 537

QY 463 TGAGCAGCAGCCTCTGACCGTGGCGGAGAGCTGACGCGGACTTCTGATCTGAT 522
DB 463 TGAGCAGCAGCCTCTGACCGTGGCGGAGAGCTGACGCGGACTTCTGATCTGAT 522

QY 538 CCGCGTGAAGTGAAGCGCGGAGCGCTCTTCTTCTTCTGAGTGAAGGCGAGTCTTA 597
DB 538 CCGCGTGAAGTGAAGCGCGGAGCGCTCTTCTTCTTCTGAGTGAAGGCGAGTCTTA 597

QY 523 CCGCGTGAAGTGAAGCGCGGAGCGCTCTTCTTCTTCTGAGTGAAGGCGAGTCTTA 582
DB 523 CCGCGTGAAGTGAAGCGCGGAGCGCTCTTCTTCTTCTGAGTGAAGGCGAGTCTTA 582

QY 598 CTTCACCTCCATATTTCTGAGAGAGCAGCGGAGTCAATTCATGAGTGGCGGCT 657
DB 598 CTTCACCTCCATATTTCTGAGAGAGCAGCGGAGTCAATTCATGAGTGGCGGCT 657

QY 583 CTTCACCTCCATATTTCTGAGAGAGCAGCGGAGTCAATTCATGAGTGGCGGCT 642
DB 583 CTTCACCTCCATATTTCTGAGAGAGCAGCGGAGTCAATTCATGAGTGGCGGCT 642

QY 658 CTGAGTCAAGTGAAGAGCAGTGTGAGAGCAGTCACTACGCGGAGTCAAGCGGAGT 717
DB 658 CTGAGTCAAGTGAAGAGCAGTGTGAGAGCAGTCACTACGCGGAGTCAAGCGGAGT 717

QY 643 CTGAGTCAAGTGAAGAGCAGTGTGAGAGCAGTCACTACGCGGAGTCAAGCGGAGT 702
DB 643 CTGAGTCAAGTGAAGAGCAGTGTGAGAGCAGTCACTACGCGGAGTCAAGCGGAGT 702

QY 718 GCCCACTGTTGCGGAGTGAACCAAGAGCGGTAATGCGCGGAGGAGGAGCAAGGTGT 777
DB 718 GCCCACTGTTGCGGAGTGAACCAAGAGCGGTAATGCGCGGAGGAGGAGCAAGGTGT 777

QY 703 GCCCACTGTTGCGGAGTGAACCAAGAGCGGTAATGCGCGGAGGAGGAGCAAGGTGT 762
DB 703 GCCCACTGTTGCGGAGTGAACCAAGAGCGGTAATGCGCGGAGGAGGAGCAAGGTGT 762

QY 778 GAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 837
DB 778 GAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 837

QY 763 GAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 822
DB 763 GAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 822

QY 838 GTGAGTCAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 897
DB 838 GTGAGTCAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 897

QY 823 GTGAGTCAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 882
DB 823 GTGAGTCAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 882

QY 898 CGTGGCGAGCAGTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 957
DB 898 CGTGGCGAGCAGTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 957

QY 883 CGTGGCGAGCAGTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 942
DB 883 CGTGGCGAGCAGTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 942

QY 958 CCGCAATTTGAGAGCGGCTGCTGATCCGCTCAAAAGTCCGCGGAGTGAAGTGAAGTGAAGT 1017
DB 958 CCGCAATTTGAGAGCGGCTGCTGATCCGCTCAAAAGTCCGCGGAGTGAAGTGAAGTGAAGT 1017

QY 943 CCGCAATTTGAGAGCGGCTGCTGATCCGCTCAAAAGTCCGCGGAGTGAAGTGAAGTGAAGT 1002
DB 943 CCGCAATTTGAGAGCGGCTGCTGATCCGCTCAAAAGTCCGCGGAGTGAAGTGAAGTGAAGT 1002

QY 1018 CCGGTTGCTGTTGAGAGCGGAGTCACTCCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1077
DB 1018 CCGGTTGCTGTTGAGAGCGGAGTCACTCCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1077

QY 1003 CCGGTTGCTGTTGAGAGCGGAGTCACTCCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1062
DB 1003 CCGGTTGCTGTTGAGAGCGGAGTCACTCCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1062

QY 1078 CTGTAATCTCTCTTGAAGCGCGCTTCAACTCGGAGTCCAGATCAAGCGGCTCTGGA 1137
DB 1063 CTGTAATCTCTCTTGAAGCGCGCTTCAACTCGGAGTCCAGATCAAGCGGCTCTGGA 1122

QY 1138 CAATGCGGAGAGTCAATGAGCGTGAACCAATCCGCGGAGTCACTAGTGAAGCGGCG 1197
DB 1123 CAATGCGGAGAGTCAATGAGCGTGAACCAATCCGCGGAGTCACTAGTGAAGCGGCG 1182

QY 1198 TCGCGCGGAGCATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGCTTACGA 1257
DB 1183 TCGCGCGGAGCATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGCTTACGA 1242

QY 1258 ACCTGCTAAGCGCGGCTCTCTTCTGAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1317
DB 1243 ACCTGCTAAGCGCGGCTCTCTTCTGAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1302

QY 1318 CACCATCTGAGTGTGGGCGGCGCACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1377
DB 1303 CACCATCTGAGTGTGGGCGGCGCACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1362

QY 1378 CAGCGCGGAGCCTTCTACGAGTGGCGTCACTGAGCAATGAGAACCTTCTCAATGA 1437
DB 1363 CAGCGCGGAGCCTTCTACGAGTGGCGTCACTGAGCAATGAGAACCTTCTCAATGA 1422

QY 1438 TTGCGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1497
DB 1423 TTGCGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1482

QY 1498 GTCCGCAAGAGGAGTCTCGGCGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1557
DB 1483 GTCCGCAAGAGGAGTCTCGGCGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1542

QY 1558 CGCCAGATGAGCAGCAGCAGCAGCAGTGAAGTCACTTCAACCAACCATGAGTGGCGGAT 1617
DB 1543 CGCCAGATGAGCAGCAGCAGCAGCAGTGAAGTCACTTCAACCAACCATGAGTGGCGGAT 1602

QY 1618 TGAAGGAGAGCAGACCACTTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1677
DB 1603 TGAAGGAGAGCAGACCACTTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1662

QY 1678 TGAAGTCAACCGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1737
DB 1663 TGAAGTCAACCGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1722

QY 1738 GTTCTTCCGTTGAGCAGAGTCAAGTGAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1797
DB 1723 GTTCTTCCGTTGAGCAGAGTCAAGTGAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1782

QY 1798 GAGTGAAGCAGCAAAAGAGCAGCAGCAGCAGTGAACGAGTGAAGTGAAGTGAAGTGAAGT 1857
DB 1783 GAGTGAAGCAGCAAAAGAGCAGCAGCAGCAGTGAACGAGTGAAGTGAAGTGAAGTGAAGT 1842

QY 1858 CTGCGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1917
DB 1843 CTGCGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1902

QY 1918 CGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1977
DB 1903 CGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1962

QY 1978 CAAAGATGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2037
DB 1963 CAAAGATGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2022

QY 2038 CTGTTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2097
DB 2023 CTGTTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2082

QY 2098 TCGGAAATCTTGTGCAATTCATCTGCTGAGGAGGAGGAGTCCGAGATTTGCTGCTGAGC 2157
DB 2083 TCGGAAATCTTGTGCAATTCATCTGCTGAGGAGGAGGAGTCCGAGATTTGCTGCTGAGC 2142

QY 2158 CTGGATCTGATCAACGTGACCTGGATGACTGTTTCTGACAAATTAATGACTTAAC 2217
Db 2143 CTGGATCTGATCAACGTGATCTGGATGACTGTTTCTGACAAATTAATGACTTAAC 2202
QY 2218 CAGGATGAGCTGCGGATGTTATCTTCAAGATTTGGCTGAGACAACTCTCTGAGGCA 2277
Db 2203 CAGGATGAGCTGCGGATGTTATCTTCAAGATTTGGCTGAGACAACTCTCTGAGGCA 2262
QY 2278 TTGCGAGTGGGACTTGAACCTGGAGCCCGAAGCCCAAGCCCAAGCAAGCAAGCAAG 2337
Db 2263 TTGCGAGTGGGACTTGAACCTGGAGCCCGAAGCCCAAGCCCAAGCAAGCAAGCAAG 2322
QY 2338 AGGACGAGCGCGGGGTGTTGCTTCTGCTGCTACAGTACCTGAGACCTTCAACGAC 2397
Db 2323 AGGACGAGCGCGGGGTGTTGCTTCTGCTGCTACAGTACCTGAGACCTTCAACGAC 2382
QY 2398 TCGACAAAGGGGAGCCCTCAACGCGGAGACGACGCGCCCTTGAAGACAAAGGCT 2457
Db 2383 TCGACAAAGGGGAGCCCTCAACGCGGAGACGACGCGCCCTTGAAGACAAAGGCT 2442
QY 2458 ACGACACAGCTCTAAAGCGGGTGAACATCCGTACTGCTGATTAACGACGCGACGCG 2517
Db 2443 ACGACACAGCTCTAAAGCGGGTGAACATCCGTACTGCTGATTAACGACGCGACGCG 2502
QY 2518 AGTTTCAGAGCGCTGCTGAAGAAATACGCTTTTGGGGGCAACCTCGGCGAGCACT 2577
Db 2503 AGTTTCAGAGCGCTGCTGAAGAAATACGCTTTTGGGGGCAACCTCGGCGAGCACT 2562
QY 2578 TCGACGCGCAAGAGGGGTCTTCTGAACTCTGCTGCTGCTGAGGAGCGCTAAGCG 2637
Db 2563 TCGACGCGCAAGAGGGGTCTTCTGAACTCTGCTGCTGCTGAGGAGCGCTAAGCG 2622
QY 2638 CTCTGGAAGAAAGCTCCGCTAGAGCAGTCCGACCAAGACCTGCTCTCGGCA 2697
Db 2623 CTCTGGAAGAAAGCTCCGCTAGAGCAGTCCGACCAAGACCTGCTCTCGGCA 2682
QY 2698 TCGGCAAGACAGCGCCGAGCGCCGCTAAAGAACTCAATTTTGGTCAACTGCGCACT 2757
Db 2683 TCGGCAAGACAGCGCCGAGCGCCGCTAAAGAACTCAATTTTGGTCAACTGCGCACT 2742
QY 2758 CAGAGTCACTCCCGATCCCAACCTCTCGAGAACTCCGACCAAGCCCGCTCTGCG 2817
Db 2743 CAGAGTCACTCCCGATCCCAACCTCTCGAGAACTCCGACCAAGCCCGCTCTGCG 2802
QY 2818 GACCTACTACATGAGCTTCAAGCGGTGCGACCAATGAGCAACATTAACAGAGCGCG 2877
Db 2803 GACCTACTACATGAGCTTCAAGCGGTGCGACCAATGAGCAACATTAACAGAGCGCG 2862
QY 2878 ACGAGTGGGTAAATGCTCAGAGAAATTTGCAATTTGCTGATTCATGCTGGGCGACAG 2937
Db 2863 ACGAGTGGGTAAATGCTCAGAGAAATTTGCAATTTGCTGATTCATGCTGGGCGACAG 2922
QY 2938 TCATACCAACAGACCCGCACTGGGCTTGGCCCACTAATTAACCACTCTTAAGC 2997
Db 2923 TCATACCAACAGACCCGCACTGGGCTTGGCCCACTAATTAACCACTCTTAAGC 2982
QY 2998 AAATCTCCAGTCTTCAACGCGGGGCGACAAACCACTACTTGGCTACAGACCC 3057
Db 2983 AAATCTCCAGTCTTCAACGCGGGGCGACAAACCACTACTTGGCTACAGACCC 3042
QY 3058 CTTGGGGGATTTTGAATTTCAACAGATTCATGCTCACTTTTCAACAGCTGAGCGAC 3117
Db 3043 CTTGGGGGATTTTGAATTTCAACAGATTCATGCTCACTTTTCAACAGCTGAGCGAC 3102
QY 3118 GACTCATCAACAAATTTGGGATTTCCGCGCCAGAGCTCACTTAATCTTCAACA 3177
Db 3103 GACTCATCAACAAATTTGGGATTTCCGCGCCAGAGCTCACTTAATCTTCAACA 3162
QY 3178 TCCAAAGTCAAGAGGTCAACAGATGATGCTCAACCACTGCTAATTAACCTTACA 3237
Db 3163 TCCAAAGTCAAGAGGTCAACAGATGATGCTCAACCACTGCTAATTAACCTTACA 3222
QY 3238 GCAAGTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTAAGCTCTGCGC 3297

Db 3223 GCAAGTTCAAGTCTTCTCGGACTCGAGTACAGATGCGGTGCTGCGCTGCGC 3282
QY 3298 ACCAGGCTGCTCCCTCCGCTTCCGCGGACGTTTCAATGATTCGCAATAGGCTAC 3357
Db 3283 ACCAGGCTGCTCCCTCCGCTTCCGCGGACGTTTCAATGATTCGCAATAGGCTAC 3342
QY 3358 TGAAGCTCAACAAATGAGCAGCAGCGGTGGAAGCTTCACTTTTCACTGCTGAAATAT 3417
Db 3343 TGAAGCTCAACAAATGAGCAGCAGCGGTGGAAGCTTCACTTTTCACTGCTGAAATAT 3402
QY 3418 TCCCTTCTCAAGTCTGAGAACGCGGCAACATTTTACCTTCACTTACCTTTGAGAG 3477
Db 3403 TCCCTTCTCAAGTCTGAGAACGCGGCAACATTTTACCTTCACTTACCTTTGAGAG 3462
QY 3478 TGCCTTTCACAGAGTACAGGCAACAGCGGAGCTGAGCGGCTGATGAATCTCTCA 3537
Db 3463 TGCCTTTCACAGAGTACAGGCAACAGCGGAGCTGAGCGGCTGATGAATCTCTCA 3522
QY 3538 TCGACCAATACCTGATTTTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAACA 3597
Db 3523 TCGACCAATACCTGATTTTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAACA 3582
QY 3598 AGGACTTGTCTTTAGCGGTGCTTCACTGAGCATGTTCTTCAAGCCCAAAATCTGCG 3657
Db 3583 AGGACTTGTCTTTAGCGGTGCTTCACTGAGCATGTTCTTCAAGCCCAAAATCTGCG 3642
QY 3658 TACCTGACCTCTGATTTTACGCGACAGCGGCTTCTAAACAAAAACAGACAAACACA 3717
Db 3643 TACCTGACCTCTGATTTTACGCGACAGCGGCTTCTAAACAAAAACAGACAAACACA 3702
QY 3718 GCAATTTTACCTGAGCTGCTTCAAAATATATACCTCAATGGGCGTGAATCATCA 3777
Db 3703 GCAATTTTACCTGAGCTGCTTCAAAATATATACCTCAATGGGCGTGAATCATCA 3762
QY 3778 ACCCTGCACTGCTATGCTTCAACAAAGACAGACAAAGTCTTTCCATGAGCG 3837
Db 3763 ACCCTGCACTGCTATGCTTCAACAAAGACAGACAAAGTCTTTCCATGAGCG 3822
QY 3838 GTGCTATGATTTTGGAAAGAGAGCGCGGAGCTTCAAACTGCTGATGACATGTCA 3897
Db 3823 GTGCTATGATTTTGGAAAGAGAGCGCGGAGCTTCAAACTGCTGATGACATGTCA 3882
QY 3898 TGAATTCAGACGAAGAGAAATTAAGCCACTAACCTGTGGCCACGGAAGATTTGGGA 3957
Db 3883 TGAATTCAGACGAAGAGAAATTAAGCCACTAACCTGTGGCCACGGAAGATTTGGGA 3942
QY 3958 CCGTGCATGCTATTTCCAGAGACAGACCACTTGGACCGGAGATGTGATGCTA 4017
Db 3943 CCGTGCATGCTATTTCCAGAGACAGACCACTTGGACCGGAGATGTGATGCTA 4002
QY 4018 TGGGAGCATTAACCTGGGATGTTGGCAAGATGAGACGTTACTGAGAGGTCCCATTT 4077
Db 4003 TGGGAGCATTAACCTGGGATGTTGGCAAGATGAGACGTTACTGAGAGGTCCCATTT 4062
QY 4078 GGGCCAAATTTCTCAACAGATGACATTTTCAACCGGTCTCTTATGAGGCGCTTGG 4137
Db 4063 GGGCCAAATTTCTCAACAGATGACATTTTCAACCGGTCTCTTATGAGGCGCTTGG 4122
QY 4138 GACTCAAGAACCGGCTCTCAATCTCTCAATCAAAACAGGCTGTTCTGCAATCTTC 4197
Db 4123 GACTCAAGAACCGGCTCTCAATCTCTCAATCAAAACAGGCTGTTCTGCAATCTTC 4182
QY 4198 CGGCGGAGTTTCAAGTCAAAAGTTTGTCTTCAATCAACCAATCTCAACAGAGCAAG 4257
Db 4183 CGGCGGAGTTTCAAGTCAAAAGTTTGTCTTCAATCAACCAATCTCAACAGAGCAAG 4242
QY 4258 TGAAGTGAATTTGAATGAGCTCAGAAAGAAACAGCAAGCGCTGGAATCCGGAAG 4317
Db 4243 TGAAGTGAATTTGAATGAGCTCAGAAAGAAACAGCAAGCGCTGGAATCCGGAAG 4302
QY 4318 TGAAGTGAATTTGAATGAGCTCAGAAATCTGCAAGTTTATCTGAGACAAAG 4377

D	b		4303	TGCAGTATACACTAATAATGCAAAATCTGCCAACGGTTGATTACCTGTGGACAACANTG	4362
O	y	4378	GACCTTATATCGAGCCTGCCCCCATTTGGCACCCGTTACCCTTAACCGCTCCCTGTAATTAC	4437	
D	b	4363	GACCTTATATCGAGCCTGCCCCCATTTGGCACCCGTTACCCTTAACCGCTCCCTGTAATTG	4422	
O	y	4438	GTCCTTATATCAATTAACCGGTTGATTCGTTTGAGTTGAACCTTGGCTCTGCTGCTCTTA	4497	
D	b	4423	GTGTTTATATCAATTAACCGGTTTATTCGTGTCAGTTGAACCTTGGCTCTGCTGCTCTTA	4482	
O	y	4498	TCTTATC-GGTTACCATGATTATAGCTTACACATTAACTGCTGGTTGCGTTGCGGATA	4556	
D	b	4483	TCTTATCTGTGTCACCATAGCAACCGGTTATACATTAACTGCTGTTGTTGCGTTGCGGA--	4540	
O	y	4557	AAAGACTTAAGTCATCGGCTTACCCTTAGTGATGAGATTGCCACTCCTCTCTCTGCGCGC	4616	
D	b	4541	-----ATAACCCCTGTGATGAGATTGCGCACTCCCTCTATAGCGCGC	4581	
O	y	4617	TGCGTCGCTGCGTGGGCGCTCGCGACCAAAGGTCGCGACAGCGGACGACTGCTCTGCGC	4676	
D	b	4582	TGCGTCGCTGCGTGGGCGCGGAGAGCAAGCTCTGCGCTGCGGACCTTTGTCGCGA	4641	
O	y	4677	GGCCCCACCGAGCGAGCGCGCGACAGAGGAGTGGGCAAA	4718	
D	b	4642	GGCCCCACCGAGCGAGCGCGCATAGAGGAGTGGGCAAA	4683	
R	E	S	AVU48704	4726 bp DNA linear VRL 15-JUN-1996	
L	O	C	AVU48704		
D	E	F	Adeno-associated virus 3 nonstructural protein and capsid protein genes, complete cds, and complete genome.		
A	C	S	U48704.1 GI:1408467		
V	E	R	Adeno-associated virus 3 strain=JH.		
R	E	F	adeno-associated virus 3		
A	H	S	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.		
T	I	L	1 (bases 1 to 4726)		
R	E	F	Muramatsu,S., Mizukami,H., Young,N.S. and Brown,K.E.		
A	H	S	Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3		
T	I	L	Virology 221 (1), 208-217 (1996)		
J	O	R	96266430		
M	E	D	8661429		
P	U	B	2 (bases 1 to 4726)		
R	E	F	Muramatsu,S. and Brown,K.E.		
A	H	S	Direct Submission		
T	I	L	Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,		
J	O	R	NHLBI/NIH, 9000 Rockville, MD 20892, USA		
F	E	A	location/Qualifiers		
S	O	U	1..4726		
			/organism="adeno-associated virus 3"		
			/strain="JH"		
			/db_xref="taxon:46350"		
			318..2192		
			/codon_start=1		
			/product="nonstructural protein"		
			/protein_id="AAC5048.1"		
			/db_xref="GI:1408468"		
			/translation="MPGEYELVLKVPDDLDERLFGISNPFVNVAWEKEMDVPPDSMDNDP NLLEOAPLVEAKLQRELFVEMRRVSAPALFPVOFEKETVFHLHLIETIGVSKS MVRVYSQIEKLVTIRIRGVGPOLPMFPAVTKRNGAGGNVDVCYINLLPERE TOPELOAMWTMDQYLASCLNLAEKRKLVAOHLYHSOTOGKONENPNSDAIVIRS KTSARVELWGLMLVDRGITSEKMWIOEDQASVISFNMASSRSOIKALKNASIMSL TKTPADTVGSNPBPDITKNRIYQILELNGTDPOYAASFGLMKQFKGRNTIWLWG PATYGKTNIAIALHAAPVFCVMWNTEPNPCVDMVLMWEEGKMPLAVNSAAFA ILGSKXRVDDCKSSAQIEPTPIVITSNTMCADVIGNSTTFEHQDLPDRMEFEL TRLRDHPGKTQEKQEVDFRWMSDHDVDAHEFYVRKGAKKRPASNDADVSPKKEE CTSIAOPTTSDAABADYADRYOKKSCHVGMINMLPCKTCERMANOISNYCFTHGOR DCGRCFPMBSQSIVSVKKTTYQLCTPHIIILBRAPIELACSADLANVLDLDCVSGQ		
C	D	S	2209..4419		

			/codon_start=1	
			/product="capsid protein"	
			/protein_id="PAC55049.1"	
			/db_xref="GI:1408469"	
			/translation="MADQGYLPMLIEDNLSEGIEMWALKGVPOPPAKNOHODNRRGKLVPEYKYLGGPGLDGEVPMENADALHEDKRAYDQOLKAGNDPYLKNHADAPEQDRIEDPSFGGKILGRAVFOAKRRIILEPIGLVEBAKTA PGKKWAGDS RQPSDSSGGV KSKGQPARKRLNFGQTDSESVDPDQLEGPPAPLPSLSGNTMASGGGAPMADNBECA DVGNSGNNWCDSQWLDKRVITITSTYMLPYTNHLVYQISSQASANDNHYFGYS TPGVDPNRFHCHFSRQMDRIINNMGWRPKLSKPLNLI QVRGTQNDGTYTILAN NLSITVPEFTDSEYQLEPYLVSAGHQCPLPEPDVFWPYGLTYLNLNGQVARSSEF YCLIEYFESOMLRTGNNGFQFTFEDVDFHVSVAHSGSDLDLMDPLIDQVYLYLRNTOG YTLGNTNOSRLPFSQAGPOSMQIQAHWMLGPGCTROORLSKTDANNNSNPWPTAAKSK YHLNGRDSLVNPPBPAMASHKQDSEKFPFMHGNLILFGKEGTHIASNAELDWNVITDEET RTTNPAVTEYQIVANNLQSSNTAPITTYGHQALPGWMOQDVLQDPIWAKIIPH TDGHFHPSPLMGFGKLHPPQPIQMIKTPVPAPNPPTFSPAKFSFITYQSTGQVSEI MEHLQENSKRNMPLEIYQSYNYSKSNVDFVTDNNGVSEPRPIGTRYILRNLI"	
BASE COUNT	1200 a	1295 c	1251 g	980 t
ORIGIN				
Query Match	68.1%	Score 3211.6;	DB 14;	Length 4726;
Best Local Similarity	81.4%;	Pred. No. 0;		
Matches 3865;	Conservative 0;	Mismatches 829;	Indels 56;	Gaps 10;
QY	1	TTGGCCCACTCCCTCTCTGCGGCGCTCGTCTCGTGGGGCTTCGGACCAAGAGTCCGC	60	
Db	1	TTGGCCCACTCCCTCTTAAGCCACTCGCTCGCTCGTGGGGCTTCGGACCAAGAGTCCGC	60	
QY	61	AGACGGACGAGCTGTGCTGCGCGGGCCCAACCGACGGACGGAGCGCGACAGAGGAGTG	120	
Db	61	AGACGGACGTGCTTTTGACGCTCCGCGCCCAACCGACGGAGGGAGTGCCCATAGAGGAGTG	120	
QY	121	GGCAACTCCATCACTAGAGGGTAATCGGAAAGCGCTCCACGCTGCCGCTCAGCGCTGA	180	
Db	121	GCAACTCCATCACTAAGGTAATGGCAGTAGACGTAACGTAAGCGGAAGCGCGCAAGACCA	180	
QY	181	CGTAATTAAGCTCATTAGGGAGTGTGCTCTGTAATTAGCTGTCACTGATGCTTTGGCA	240	
Db	181	CGCTTACCA-----GCTCGTCAGCAGTCAAGGTGA-CCCTTTTGGGAC	222	
QY	241	ATTTTGGACACCAAGTGGCAATTTAGGGTAATATATGCGCGAGTGAAGG-AGACGAGATCT	299	
Db	223	AGTTTGGACACCAAGTGGCGCGCTGAGGGTAATATCTGAGTGAACGAAACGAGAGCT	282	
QY	300	CCATTTTGACCGCGGAATTTGAACAAGACAGACGCAATGCGGGCTTCTACAGATCGTGA	359	
Db	283	CCATTTTGACCGCGAATTTGAAAGACAGACAGCAATCCGGGGTCTTACAGATTTGTCC	342	
QY	360	TCAAAGTGCAGAGCACTTGAAGACGACCTGCGGGCAATTTGACTCGTTTGTGAGCT	419	
Db	343	TGAAGGTCCGAGTGACTTGAAGAGGCGCTGCGGGCAATTTCTAATCTCGTTTGTAACT	402	
QY	420	GGGTGGCGGAGGAATTGGAGCTGCCCCCGGATTTGACATGATCTGAATCTGATTG	479	
Db	403	GGGTGGCGGAGGAATGGAAGTGAAGTGCAGCTTGTGACATGGAATCCGAATCTGATTG	462	
QY	480	AGCAGGACCCCTGACCGTGAAGGAGTGAAGCGACGCACTTCCCTGATCAATGGGCGCC	539	
Db	463	AGCAGGACCCCTGACCGTGAAGGAGTGAAGCGAGTTCACGCGAGATTCCTGATGAGTGGCGCC	522	
QY	540	GCGTGAAGAGCCCGCGAGGCGCTCTCTTTTGTTCAGTTCAGTAAGAGGCGAGTCTACT	599	
Db	523	GCGTGAAGTAAAGCCCGCGAGGCGCTCTTTTGTTCAGTTCGAAAAGGGGAGACTACT	582	
QY	600	TTCACCTTCATATTTGTGTGGAACCAACGGGGGTCAATTCATGTGTCTGGGCGCTTCC	659	
Db	583	TTCACCTTCACGTGTGATTAAGACCAATCGGGGTCAATTCATGTGTGTGGCGCGCTACG	642	
QY	660	TGAGTCAAGTTAAGGACAAGCTGATGACCAATCTACCGCGGATATGAGCCGACCTTGC	719	
Db	643	TGAGCAGATTAAGAAGAGACTGTGACCCGCACTTACCGCGGGGTGAGCCCGACGTTTC	702	
QY	720	CCAACGTGTTCCGCGTGACCAAGACGGTAATGAGCGCGGAGGGGAGAACAGAGTGTGG	779	

Db 703 CGAAGTGGTTCGCGGTGACCAAAAGCGAATGCGCGGCGGGAACAAGTGGTGG 762
Qy 780 ACGAGTGTATATCCCACTACTCTGCGCAAGACTCAGCCGAGTGTGAGTGGCGT 839
Db 763 ACGACTCTATCATCCCACTACTCTGCGCAAGACTCAGCCGAGTGTGAGTGGCGT 822
Qy 840 GGAATTAATAGGAGATATATAGCGCTGTGAACTGTGGCGAGGCGAAAGCGCTGG 899
Db 823 GGAATTAATAGGAGATATATAGCGCTGTGAACTGTGGCGAGGCGAAAGCGCTGG 882
Qy 900 TGGCGAGCACTGACCCAGCTGACGAGCAAGCAAGAGAGAGAGAGATGTAACC 959
Db 883 TGGCGAGCACTGACCCAGCTGACGAGCAAGCAAGAGAGAGAGAGATGTAACC 942
Qy 960 CCAATTTCTGAGCGCGCTGTATCCGCTCAAAAACCTCGCGGCTCAATGAGTGGTGG 1019
Db 943 CCAATTTCTGAGCGCGCTGTATCCGCTCAAAAACCTCGCGGCTCAATGAGTGGTGG 1002
Qy 1020 GGTGGCTGTGGAGCGGGGGCATCACTCCGAGAGAGTGGATCGAGAGGACCAAGGCT 1079
Db 1003 GGTGGCTGTGGAGCGGGGGCATCACTCCGAGAGAGTGGATCGAGAGGACCAAGGCT 1062
Qy 1080 CGTACATCTCTCTCAAGCGCGCTTCAACTCGCGGTCCAGATCAAGCGCTCTGACA 1139
Db 1063 CGTACATCTCTCTCAAGCGCGCTTCAACTCGCGGTCCAGATCAAGCGCGCTCTGACA 1122
Qy 1140 ATGCGCGAGATATAGCGCTGTACCAATCCGCGCGGCTCACTGTGAGGCGCGCTC 1199
Db 1123 ATGCGCTCAAGATATAGCGCTGTACCAAGCGCGCTCCGAGATCACTGTGAGGCGAGAAC 1182
Qy 1200 CGCGCGGAGCATTAACCAACCGCATCTACCGGATCTGTGAGCTGAGCGCTACGAGAC 1259
Db 1183 CGCGCGGAGCATTAACCAACCGCATCTACCGGATCTGTGAGCTGAGCGGCTACGAGAC 1242
Qy 1260 CTGCGTACGCGCGCTCGCTCTTCTCGGCTGAGGCGCAAAAAGTTCGAGAGCGCAACA 1319
Db 1243 CGCGTACGCGCGCTCGCTCTTCTCGGCTGAGGCGCAAAAAGTTCGAGAGCGCAACA 1302
Qy 1320 CCATCTGCTGTGTGGGCGCGGCGCAACAGGGGCAAGACCAATGTGGCGAAGCTCCGCC 1379
Db 1303 CCATCTGCTGTGTGGGCGCGGCGCAACAGGGGCAAGACCAATGTGGCGAAGCTCCGCC 1362
Qy 1380 ACGCGTGGCCCTTCTACGCGCTGCGTCAATGAGCAATGAGACCTTCCCTCAATGAT 1439
Db 1363 ACGCGTGGCCCTTCTACGCGCTGCGTCAATGAGCAATGAGACCTTCCCTCAATGAT 1422
Qy 1440 GCGTCAACAAGATGTATCTGTGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGTGAGT 1499
Db 1423 GCGTCAACAAGATGTATCTGTGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGTGAGT 1482
Qy 1500 CCGCGAAGGCGATCTCGGCGGCGAGCAAGGTGCGCGTGTGACCAAAAAGTGTGAGTCCG 1559
Db 1483 GCGCGAAGGCGATCTCGGCGGCGAGCAAGGTGCGCGTGTGACCAAAAAGTGTGAGTCCG 1542
Qy 1560 CCGAGATCGACCCCAAGCGCGTGTGATCACTTCAACCAACCAATGTGCGCGTGTGAT 1619
Db 1543 CCGAGATCGACCCCAAGCGCGTGTGATCACTTCAACCAACCAATGTGCGCGTGTGAT 1602
Qy 1620 ACGGGAACAGCAACCTTGTGAGACCAAGCGCGTGTGAGACCGGATGTTCAATTTG 1679
Db 1603 ACGGGAACAGCAACCTTGTGAGACCAAGCGCGTGTGAGACCGGATGTTCAATTTG 1662
Qy 1680 AACTCAACCGCGCTGTGAGATGATCTTGTGCAAGGTGACCAAGAGAGTCAAAAGT 1739
Db 1663 AACTCAACCGCGCTGTGAGATGATCTTGTGCAAGGTGACCAAGAGAGTCAAAAGT 1722
Qy 1740 TCTTCGCTGGGCGCAGATCACTGACCGAGGTGTGCGCATGATTTCACTGAGAAAG 1799
Db 1723 TTTTTCGCTGGGCTTTCGATCACTGACGTGTGCTCATGATTTTCACTGAGAAAG 1782
Qy 1800 GTGAGCGCAACAAAAGACCGCGCCCGATGACGCGGATTAAGAGAGCCCAAGCGGCGCT 1859

Db 1783 GTGAGCTAAGAAACCCCGCTCCATGACGCGGATGTAAGCGAGCAACCGGAGT 1842
Qy 1860 GCGCTCAGTGGGATTCATGACGTGACAGCGCGAGAGAGTCCGCTGAGATTTGGCG 1919
Db 1843 GCACTCAGTGGGATTCATGACGTGACAGCGCGAGAGAGTCCGCTGAGATTTGGCG 1899
Qy 1920 ACAGTACCAAAACAATGTCTGTGACGCGGCGATGCTTCAATGTGCTGCTGCA 1979
Db 1900 ACAGTACCAAAACAATGTCTGTGACGCGGCGATGCTTCAATGTGCTGCTGCA 1959
Qy 1980 AGCATGCGAGAGATGATCAGATTTCAATTTGCTTCAACGCAAGGAGAGAGACT 2039
Db 1960 AGCATGCGAGAGATGATCAGATTTCAATTTGCTTCAATTTGCTTCAACGCAAGGAGACT 2019
Qy 2040 GTTCAGAGTCTCCCGCGGTGCTGAGATCTGAC-----CGGTGTGAGAAAGAG 2093
Db 2020 GTGGGAGATCTTCCCTGAGATGTCAGATCTAACCCGTTTCTGTGTGTAAGAGAG 2079
Qy 2094 CGTATCGAAGACTCTGTGCAATTCATCATCTGTGAGGCGGAGCTCCGAGATTTGCT 2153
Db 2080 CTATCAGAAACTGTGTCAATTCATCATCTGTGAGGAGGAGCAAGGAGATTTGCT 2139
Qy 2154 CGGCTGCGATCTGTGACGTGAGCTGAGATCTGTGCTTCTGAGCAATTAATGACT 2213
Db 2140 CGGCTGCGATCTGTGACGTGAGCTGAGATCTGTGCTTCTGAGCAATTAATGACT 2199
Qy 2214 AAACAGATATGCTGCTCCATGATTTCTTCCAGATTTGCTGAGAGCAACCTCTGAG 2273
Db 2200 AAACAGATATGCTGCTCCATGATTTCTTCCAGATTTGCTGAGAGCAACCTCTGAG 2259
Qy 2274 GGCATTCGAGTGTGGAGCTTGAACCTGAGACCCCGAAGCCCAAGCCACAGCA 2333
Db 2260 GGCATTCGAGTGTGGAGCTTGAACCTGAGACCCCGAAGCCCGAAGCCACAGCA 2319
Qy 2334 AAGCAGAGCAGCGCGGCTGTGCTTCTGCTGCTCAAGTACTTCCGACCTTCAAC 2393
Db 2320 CACGAGAGCAACCGTGGGCTTGTGTCTTCCGAGTTCAAAATCTTCCGACCCGCTAAC 2379
Qy 2394 GGACTGAGCAAGGAGGAGCCCGTCAAGCGGCGAGCGAGCGGCTCTGAGCAGCAAG 2453
Db 2380 GGACTGAGCAAGGAGGAGCCCGTCAAGCGGCGAGCGAGCGGCTCTGAGCAGCAAG 2439
Qy 2454 GCTTACGACGAGCTCAAAAGCGGCTGACATCTGCTGAGTATTAACAGCGGAC 2513
Db 2440 GCTTACGACGAGCTCAAAAGCGGCTGACATCTGCTGAGTATTAACAGCGGAC 2499
Qy 2514 GCGGATTTCAAGAGCTGTGCAAGATGATCTTTTGGGGCAACCTCGGCGAGCA 2573
Db 2500 GCGGATTTCAAGAGCTGTGCAAGATGATCTTTTGGGGCAACCTCGGCGAGCA 2559
Qy 2574 GTCCTCAGGCGCAAGAGCGGCTTCTGAACTCTCGGCTGTGAGAGAGCGCTAAG 2633
Db 2560 GTCCTCAGGCGCAAGAGCGGCTTCTGAACTCTCGGCTGTGAGAGAGCGCTAAG 2619
Qy 2634 ACGGCTCTGGAAGAAAGCTCGGTAGAGCACTGACCAAGAGCGAGCTCTCTCG 2693
Db 2620 ACGGCTCTGGAAGAAAGCGGCTGTAGATCACTCTCAGAGAACCGAGCTCATCT 2679
Qy 2694 GGCATGCGCAAGACAGCGCAGAGCGGCTTAAGAGAGCTCAATTTGTGAGCTGCG 2753
Db 2680 GGCATGCGCAAGACAGCGCAGAGCGGCTTAAGAGAGAGCTCAATTTGTGAGCTGCG 2739
Qy 2754 GACTCAGAGTCACTCCCGATCCAGAACTCTGAGAGAACCTCAGAGAACCCCGCTGCT 2813
Db 2740 GACTCAGAGTCACTCCCGATCCAGAACTCTGAGAGAACCTCAGAGAACCCCGATGCT 2799
Qy 2814 GTGAGCTTAATCAATGCTTTCAGCGGTGTGCGCACTAATGCGAGCAATTAAGAGC 2873
Db 2800 TTGGGATCTTAATCAATGCTTTCAGCGGTGTGCGCACTAATGCGAGCAATTAAGAGC 2859
Qy 2874 GCGGAGAGAGTGGGTATGCTTCAAGAAATTTGCAATTTGCAATGCTGCGGCGAC 2933
Db 2860 GCGGAGAGAGTGGGTATTTCTCAGAGAAATTTGCAATTTGCAATGCTGCGGCGAC 2919

[illegible]

QY	4011	CATCCATGGAGACATTAACCTGGCATGATGTCGGCAATATGAGACGTATCCCTGCACAGGT	4070
Dp	3997	AATATATCAGGGGGGCTTAACCTGCGATGATGTGGCAAAATGTGACGTATCTTCAAGGA	4056
QY	4071	CCCATTTGGGCGCAAAATTCCTCCACACAGATGACACTTTGACCCGCTCTCTTATGAGC	4130
Dp	4057	CCATCTCTGGGCAAAAGATTCTCTCACACGAGATGGACATTTTATCTTCTCTGATGGGA	4116
QY	4131	GGCTTTGGACTCAAGAAACCGGCTCTCCAGATCTCATCAAAAACAGCCTGTCTCTGCG	4190
Dp	4117	GGCTTTGGACTGAAACATCCGCGCTCTCAAAATCATGATCAAAAATATCTCCGGTACCGGCA	4176
QY	4191	AATCCCTCGGCGAGGTTTACGATCAAAAGTTTGCTCATTCATCACCAATATCTCCACA	4250
Dp	4177	AATCTCTCGAGACTTTCAGCCCGGCGCAAGTTTGCTCATTTATCACTCACTATCTCACT	4236
QY	4251	GGACAACTGAGTGTGAAATTTGATGGAGCTCGAGAAAGAAAACAGCAAGCGCTGGAAT	4310
Dp	4237	GGACAGGTTCAGCGTGGAAATTTGATGGAGCTCACAGAAAGAAAACAGCAAAACGTTGGAAT	4296
QY	4311	CCCGAAGTGCATACATCCATTTATGCAAAATCTGCCAAGCTTGATTTTACTGTGGAC	4370
Dp	4297	CCAGAGATTCAGTACACTTCCAACTACACAAAGCTGTATATGTGGACTTTACTGTAGAC	4356
QY	4371	AACAATGACCTTATATCTGAGCCTCGGCGCATTTGGACCCGTTACCTTACCCGCGCCCTG	4430
Dp	4357	ACTNATGCTGTTTATATGTAAGAACCTCGGCTATTTGAAACCCGGTATCTCACACGAACTTG	4416
QY	4431	TAATTAAGTGTATATCAATTAACCCGATTTGATTCGTTCAAGTTGAACTTTGG---TCTCC	4486
Dp	4417	TGAATCTGGTATATCAATTAACCGTTTATTTGTTCAGTGAACCTTTGGCTCTGTGTC	4476
QY	4487	TGTCCTTCTTATCTTATCGGTTACATGATGTTATAGCTTACACATTACTG-----CTTG	4540
Dp	4477	ACTTCCTTATCTTATCTTGTGTTTCCATGAGCTACGCTAATATAGACGCGGCTGCGAGC	4536
QY	4541	GTTGCGCTTGCGCATAAAGC-----TTAAGTCATCGGGTTACCCCTAGTGA	4588
Dp	4537	CTTGCGCTTGCGGTTTACAACCTGCTGGTTAATATTTTAACTCTCGCCATATCTTTATGTA	4596
QY	4589	TGAGATTGCCACTCCCTCTCTGCGGCTGCTGCTGCTGCGGTGGGGGCTGAGGAAAG	4648
Dp	4597	TGAGATTGGGCACTCCCTCTATGAGCACTGCTGCTGCTGCGGTGGGGGCTGGGCGAACAAAG	4656
QY	4649	TTCGCAAGCGGACAGGCTTCTGCTCTGCTGCGGCCCAACGAGCGAGCGAGGAGG	4708
Dp	4657	TTCGCAAGCGGAGCGTCTTGGACAGTCCGGGCCCAACGAGCGAGCGAGTGCATAGAG	4716
QY	4709	GAGTGGGCAA 4718	
Dp	4717	GAGTGGGCAA 4726	

RESULT 4	AF028705	AF028705	4722 bp	DNA	linear	VRL 12-JAN-1998
LOCUS						
DEFINITION	Adeno-associated virus 3B, complete genome.					
ACCESSION	AF028705					
VERSION	AF028705.1	GI:2766608				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						

JOURNAL Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720,
Seattle, WA 98195, USA
FEATURES Location/Qualifiers
Source .

CDS

1..4722
/organism="adeno-associated virus 3B"
/db_xref="taxon:68742"
/note="AAV3B"
317..2191
/codon_start=1
/product="nonstructural protein Reg78"
/protein_id="AA895451.1"
/db_xref="GI:2766609"
/translation="MGEFYELVKVPSDLDELHGLSINSFVNVAEKEMELPPSDMD
PNLEIOLALVLAKELOREFLVEMRVSKAEALFVQEKGETYFLHLVLIETGVK
MVGKYSOIKEKLVIRYRGVPEOLPMFAVTKRAGAGKVVDDCYIPYLLPK
TOPELOMAMTNDVOYLACINLAERKLVNQHLYVSGOEOKEKNONPRSDA.PVRS
KTSARIMELVGLVDRIITSSEKMDIDQASYISFNASVRSQIKALDNASKINSL
TKTAPYLVGSNPEDITTKRIYQILELNGYDQYASVAVYGAQKFKGRNTIWLFG
PATGKTNIAEALAHAVPYGCVMNTNENPNDKVMITWMEBGMTAKVESAKA
ILGSKVRVDQKSSAOIEPTPIVITSNTMCAVIDGNTSTFEHOPLDDRMKEFL
TRLDHDFGVTKQEVYDFPRMASHDVTVDYAHDFYVRKGGAKKRPAISNDADSPKRO
CTSIADPTSDAPADPADYADRYONKCSRHVGMMLPFCCTCERNQINSVCTHQR
DCECFPMSESQPVSVKKKTYQKLCPIHIIHLIFAPETIACADLANVLDLCVSEQ
"

CDS

2208..4418
/codon_start=1
/product="capsid protein VP1"
/protein_id="AA895452.1"
/db_xref="GI:2766610"
/translation="MAADGYLPWLELDNLSSEGLREWMALKPVOPKANOQHDNRG
LVLPGYKYLPGNGLDKGEVFNADAALEHDAVDOQLKAGNPYLKYNHADAEOE
RLQEDTSFGNGLRAVFOAKRIILEGLVEBAKTAPEKRPVQDSPOQDSSGVG
KSGKOPARKLINFQGDSESVDPDQGLPAPATSLGNTWASGGAGMADNNSGA
DGVNSGNNHCDSONLADRVITSTRTALPTYNHLYKQISOSGASNDNRYPS
TPMGYDFRFRHCFSPRDMQRLINNMGKPKLSFKLPNIQVKEYTQNDGTTTAN
NLTSYQVFTDSYQLPYVLGSAHQGLPEFPADEVMPVGYLTLLNSQAVGRSF
YCLAEPPSOMLRGTNNFOFSYFEDVDFSHSVASQSLRLNMP.LIDQYLYAVTRQG
TTEGTNOSRLIFSQAPOSMLOARNMILPCYEOOLSKTNDNNNSNPFMTASX
YHLNGRDLVLPAPAMASHKDEKFPMMGNLIFGEGETTANAELENDNMVITDEEI
RTNVPATBOYGVANNLOSNTAPTRTVDGALPGMWODRDLVYQIPWAKI.PH
TDGHPSPLMGSGFGLKHPPIQIMIKTPVPANPPTTSPSKASSTIYQSTQVSE
IEMELQENSKRNPEIQYTSNYSKSNVDFIVDVGVSSEPRIGRIYLRNL"

BASE COUNT 1204 a 1297 c 1243 g 978 t

ORIGIN

Query Match 67.9%; Score 3205; DB 14; Length 4722;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 3872; Conservative 0; Mismatches 820; Indels 55; Gaps 12;

QY 2 TGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGGCTCGGACCAAGATCCGCA 61
DB 1 TGCCCACTCCCTCTGCGGCTCGCTCGGTGGGGCTCGGACCAAGATCCGCA 60
QY 62 GACGGCAGAGCTCTGCTCTGCGGCTCGGACCGAGCGAGCGCGGAGAGAGTGG 121
DB 61 GACGGCAGAGCTCTGCTCTGCGGCTCGGACCGAGCGAGCGCGCATGAGAGTGG 120
QY 122 GCAACTCAATCACTAAGGGGTATGCGAGAGGCTCCACAGCTGCGCGCGTCAAGGCTGAC 181
DB 121 CCAACTCAATCACTAAGGGGTATGCGAGAGGCTCCACAGCGCGGAGGAGACAC 180
QY 182 GTAATTACGTCATAGGGAGTGTCTCTGATTAGCTGTCAKGTAGTCTTTTGGACA 241
DB 181 GCCTACCA-----GCTGCTCAGCAGTCAAGTGA-CCTTTTGGACA 222
QY 242 TTTTGGACACACGTCGCTATTAGGGTATATATGGCCGATGAGCG-AGCAGATCTC 300
DB 223 GTTTGGACACACGTCGCTGAGGGTATATATCTCGATGAGGGAACCGAGAGCTC 282
QY 301 CATTTGACCGGAAATTTGAGAGAGAGGACGACGCGGCGCTTCTACAGAGATCGGAT 360
DB 283 CATTTGACCGGAAATTTGAGAGAGAGGACGACGCGGCGCTTCTACAGAGATGCTCT 342

QY 361 CAAGTCCGAGGACCTTGAAGACGACGCTGCGGCGATTTCTGACTGTTTGAGCTG 420
DB 343 GAAAGTCCCGAGTGAAGTGAAGACGACGCTGCGGCGATTTCTGACTGTTTGAGCTG 402
QY 421 GGTGGCCGAGAGAAATGAGTGAAGTGGCCCGGATTTCTGACATGATCTGAATCTGATTA 480
DB 403 GGTGGCCGAGAGAAATGAGTGAAGTGGCCCGGATTTCTGACATGATCTGAATCTGATTA 462
QY 481 GCAAGCACCCTGACCGGTGGCGAGAGAGTGAAGGCGGCTTCTGCTGCAATGAGCGCG 540
DB 463 GCAAGCACCCTGACCGGTGGCGAGAGAGTGAAGGCGGCTTCTGCTGCAATGAGCGCG 522
QY 541 CGTGAATTAAGGCGCCGAGAGCGCTCTTCTTGTGTCAGTTGAGAGAGGCGAGTCTACTT 600
DB 523 CGTGAATTAAGGCGCCGAGAGCGCTCTTCTTGTGTCAGTTGAGAGAGGCGAGTCTACTT 582
QY 601 CCACTTCATATTCTGTGAGAGACCAAGGCGGTCAATCAATGATGCTGCGCGCTTCT 660
DB 583 CCACTTCATATTCTGTGAGAGACCAAGGCGGTCAATCAATGATGCTGCGCGCTTCT 642
QY 661 GAGTCAGATTAGGGAACAAGTGGTGACCAATCTACGCGGAGATGACGCCACCTGCG 720
DB 643 GAGTCAGATTAGGGAACAAGTGGTGACCAATCTACGCGGAGATGACGCCACCTGCG 702
QY 721 CAATGCTTTCGCGGTGACCAAGCGGTATGCGCGGAGAGGCGGGAACAAGTGTGGA 780
DB 703 GAACTGCTTTCGCGGTGACCAAGCGGTATGCGCGGAGAGGCGGGAACAAGTGTGGA 762
QY 781 CGAGTCTTATCCTCCCACTACTCTCTGCGCAAGACTGACGCCGAGCTGAGTGGCGTG 840
DB 763 CGAGTCTTATCCTCCCACTACTCTCTGCGCAAGACTGACGCCGAGCTGAGTGGCGTG 822
QY 841 GACTTAATGAGAGAGATATATTAAGCGCTGTTTGAACCTGCGCGAGCGCAACCGCTCG 900
DB 823 GACTTAATGAGAGAGATATATTAAGCGCTGTTTGAACCTGCGCGAGCGCTGAGTGG 882
QY 901 GCGCGAGCACTGACCCACGTCAGCGGACCAAGAGAGCAACAAGAGATCTGAACCC 960
DB 883 GCGCGAGCACTGACCCACGTCAGCGGACCAAGAGAGCAACAAGAGATCTGAACCC 942
QY 961 CAATTTGACGCGCTCTGATCCGTCACAAACCTCCGCGCTACATGAGAGTGGTCCG 1020
DB 943 CAATTTGACGCGCTCTGATCCGTCACAAACCTCCGCGCTACATGAGAGTGGTCCG 1002
QY 1021 GTGCTGTGAGACCGGGGATCACTCCGAGAAAGAGTGAATCAAGAGAGACCAAGGCTC 1080
DB 1003 GTGCTGTGAGACCGGGGATCACTCCGAGAAAGAGTGAATCAAGAGAGACCAAGGCTC 1062
QY 1081 GTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCGCGCTTGGACAA 1140
DB 1063 GTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCGCGCTTGGACAA 1122
QY 1141 TGCGGCAAGATCAAGGCGCTGACCAATCCGCGCGGCTACCTGCTGAGGCGCGCTCC 1200
DB 1123 TGCGGCAAGATCAAGGCGCTGACCAATCCGCGCGGCTACCTGCTGAGGCGCGCTCC 1182
QY 1201 GCGCGGAGCATTTAAACCAACGATCTACCGCATCTCGAGGCTGAAGCGCTACCAAC 1260
DB 1183 GCGCGGAGCATTTAAACCAACGATCTACCGCATCTCGAGGCTGAAGCGGATCAATCC 1242
QY 1261 TGCTTACGCGGCTCTGCTTCTGCGGTGGGCGCAAAAGAGTTCGGAAGCGCAAC 1320
DB 1243 GCAAGTACGCGGCTCTGCTTCTGCGGTGGGCGCAAAAGAGTTCGGAAGCGCAAC 1302
QY 1321 CATCTGCTGTTTGGGCGCGGCAACGAGGCAAGACCAATCGCGGAGGCAATCGGCCA 1380
DB 1303 CATCTGCTGTTTGGGCGCGGCAACGAGGTAACCAATCGCGGAGGCAATCGGCCA 1362
QY 1381 CGCGGTGCGCTTCTAGCGGTGCGTCAATGAGCAATGAGAACTTCCCTTCAATGATTG 1440
DB 1363 CGCGGTGCGCTTCTAGCGGTGCGTCAATGAGCAATGAGAACTTCCCTTCAATGATTG 1422
QY 1441 CGTCGACAAAGATGATCTGTGTGGAGAGAGGCAAGATGACGCGCAAGGTGCTGAGTC 1500


```
Db 1423 CGTCGACGAGTGTGATCTGCGGGAGAGGCGCAAGATGAGCGGCCAAGGTGCTGAGAG 1482
Qy 1501 CGCGAAGGCCATTTCTCGGCGGCGACGAAAGTGGCGCTGAGACCAAAATGCAATGCTGCTCCG 1560
Db 1483 CGCGAAGGCCATTTCTCGGCGGAGCAAGTGGCGCTGAGACCAAAATGCAATGCTGCTCCG 1542
Qy 1561 CGAGATCGACCCGACCCCGGTGATGCTGACCTTCACACCAACATGAGCGCGTATTTGA 1620
Db 1543 CGAGATCGAACCACTCCGCTGATGCTGACCTTCACACCAACATGAGCGCGTATTTGA 1602
Qy 1621 CGGGAACAGACCAACCTTCGAGCAGCAGACCGGTTGCGAGACCGGATGTTCAATTTGA 1680
Db 1603 CGGGAACAGACCAACCTTCGAGCAGCAGACCGGTTGCGAGACCGGATGTTCAATTTGA 1662
Qy 1681 ACTCAACCCGCGCTCTGAGAGATGATCTTTGGCAAGGTGACAAAGCAGAAAGTCAAGATT 1740
Db 1663 ACTTACCCGCGCTTTGAGCAGTATCTTTGGGAAGGTGACCAAAACAGAAAGTAAAGACTT 1722
Qy 1741 CTTCCGCTGGGCGCAGAGATCAGTGAACCGAGTGGCGCATGAGTTCTGACGTCAAGAAAGG 1800
Db 1723 TTTCCGCTGGGCTTCGATCAGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1782
Qy 1801 TGGAGCCAAACAAAGAACCCGCCCCCGATGACCGGATTAAGCAAGCAGCCAGCGGCTG 1860
Db 1783 TGGAGCTTAAGAAAGCGCCCGCTCAATGACCGGATTAAGCAAGCAGCCAGCGGCTG 1842
Qy 1861 CCCCTGATGCGCGATGATCAGCTGACGCTGACGCGGAAGAGCTCCGCTGATCTTGGCGA 1920
Db 1843 CAGTCACTTGGCGACCGACCAACGCTGACGCGGAG---AGCACCGCGGACCTACCGGA 1899
Qy 1921 CAGGTACCAAAACAAATGTTCTGTCACGCGGCGATGCTTCAAGTCTGTTCCCTGCA 1980
Db 1900 CAGGTACCAAAACAAATGTTCTGTCACGTTGGGCAAGATCTGATGCTTTTCCCTGTA 1959
Qy 1981 GACATCGAGAGATGATCAGAAATTTCAACATTTGCTTCAACGACGAGCGAGAGACTG 2040
Db 1960 AACATCGAGAGATGATCAGAAATTTCAACATTTGCTTCAACGATGCTTCAAGAGACTG 2019
Qy 2041 TTCAGATGCTTCCCGGCGCTGCAATTTCAAC-----CGGTCTCAAGAAAGAGAC 2094
Db 2020 TGGGGAATGCTTCCCTGGAATGTCAGAACTCAACCCGTTTCTGTCTCAAAAGAGAGC 2079
Qy 2095 GATCGGAAACCTGAGCATTCATCATCTGCGGCGGCGCTCCGAGATGCTGCTGCTC 2154
Db 2080 TTATCAAGAACTGTGTCATTCATCATCTGAGAAAGGCAACCGAGATGCTGCTGCTC 2139
Qy 2155 GGCCTGCACTGCTGCAACGCTGACCTGATGACTGTGTTCTGAGCAATTAATGACTTA 2214
Db 2140 GGCCTGCAATTTGGCAATGTGACTGATGACTGTGTTCTGAGCAATTAATGACTTA 2199
Qy 2215 AACCAAGTATGCTGCGCATGATGTTATCTTCAGATTTGCTGAGCAACCTCTGAGG 2274
Db 2200 AACCAAGTATGCTGCTGCAAGTTATCTTCAGATTTGCTGAGCAACCTTTCTGAGG 2259
Qy 2275 GATTTGGCAGTGTGTGAGACTTGAACCTGAGACCCCGAAGCCCAAGCCAAACCAACAGAAA 2334
Db 2260 GATTTGTGATGTGTGTGCTGTAAACCTGAGAGTCCCTCAACCCAAAGCAACCAAC 2319
Qy 2335 AGCAGAGCAGCGCGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2394
Db 2320 ACCAGAGCAACCTCGGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2379
Qy 2395 GACTGCAAGAGGGGAGCCCGCTCAACGCGCGGAGCGAGCGCGCTGCGAGCAGCAAGG 2454
Db 2380 GACTGCAAGAGGAGCGCGCTCAACGAGCGGAGCGAGCGCGCTGCGAGCAGCAAGG 2439
Qy 2455 CCAAGACCAAGCGCTCAAGCGGGTGAACAACCGTACCTGCGGTATTAACAGCCGAGG 2514
Db 2440 CTTACAGACAGCGCTCAAGCGGGTGAACAACCGTACCTCAAGTACCAACAGCCGAGG 2499
Qy 2515 CCGAGTTTCAAGAGCGCTGCAAGAAAGTACGCTTTTGGGGGCAACCTCGGCGAGCAG 2574
|||||
Db 2500 CCGAGTTTCAAGAGCGCTTCAAGAAATACGCTTTTGGGGGCAACCTTGGCAGAGCAG 2559
Qy 2575 TCTTCCAGGCGCAAGAAAGCGGGTTCGAACTCTCGCTGTGTTGAGAGAGCGCTAACA 2634
Db 2560 TCTTCCAGGCGCAAGAAAGAGATCTTGAAGCTCTTGTGTGTTGAGAGAGCAGTAAAA 2619
Qy 2635 CGGCTCTTGGAAAGAAACGTCCGATGAGCAGTCCGCAACAGAGCCAGACTCTCTCGG 2694
Db 2620 CGGCTCTTGGAAAGAAAGAGAGCGCTGATGATCACTCTCTCAGAGAACCGGACTATCTG 2679
Qy 2695 GCATGGCAAGACAGCGCAGACCGCGCTAAAGAGACTCAATTTTGTGACACTGCGG 2754
Db 2680 GTGTTGGCAAAATCGGGCAAAACAGCTCGCAAGAAAGACTAATTTGCTGACACTGCGG 2739
Qy 2755 ACTCAGAGTCACTCCCGATTCACAACTCTGAGGAACCTCAGAGAACCCCGCTGCTG 2814
Db 2740 ACTCAGAGTCACTCCCGATTCACAACTCTGAGGAACCTCAGAGAACCCCGCAAGTT 2799
Qy 2815 TGGGACCTACTCAATAGGCTTCAAGCGGTGCGGCACCAATGGCAGCAATTAAGAGCGG 2874
Db 2800 TGGGATCTAATACATAGGCTTCAAGCGGTGCGGCACCAATGGCAGCAATTAAGAGCGG 2859
Qy 2875 CCGACGAGTGGGTATATGCTCAGAGAAATGGCATTTGCGATTCACATGCTGGCGACA 2934
Db 2860 CCGATGAGTGGGTATATCTCAGAGAAATTTGCAATTTGCGATTCACATGCTGGCGACA 2919
Qy 2935 GAGTATCAACACAGACACCGGCACTTGGGCTTGGCCGCTGCAATTAACACTCTTACA 2994
Db 2920 GAGTATCAACACAGACACCAACCTGGGCTTGGCCGCTGCAATTAACCAACTCTTACA 2979
Qy 2995 AGCAAACTCCAGTGTCTCAAGCGGGGCGAGGACGACCAACCACTACTTGGCTACAGCA 3054
Db 2980 AGCAAACTCCAGGCAATCA---GGAGCTTCAAGACGACCACTACTTGGCTACAGCA 3036
Qy 3055 CCCCCTGGGGATTTTGAATTTCAACAGATTCACATCCGACTTTTCAACGAGTGACTG 3114
Db 3037 CCCCTTGGGGATTTTGAATTTCAACAGATTCACATCCGACTTTTCAACGAGTGACTG 3096
Qy 3115 AGCAGTCACTCAACCAATTTGGGATTTCCGGCCCAAGAGATCTCAACTTCAACTTTCA 3174
Db 3097 AGCAGTCACTCAACCAATTTGGGATTTCCGGCCCAAGAGATCTCAACTTCAACTTTCA 3156
Qy 3175 ACATCCAACTCAAGAGGCTGACGAGCAATGAGGCTGACCAACCACTGCTAATTAACCTTA 3234
Db 3157 ACATCCAACTTAAAGAGTCAAGCAGAGACATGAGCAGACATTAATGCAATTAACCTTA 3216
Qy 3235 CAGCAAGGTTCAAGTCTTCTGCACTGAGACTGAGTACCAAGCTTCCGTAAGTCTGAG 3294
Db 3217 CAGCAAGGTTCAAGTCTTCTGCACTGAGACTGAGTACCAAGTCTTCCGTAAGTCTGAG 3276
Qy 3295 CGCACAGGCTGCTCTCCCTCCGTTCCGGGCGGAGCTGTTCATGATTCGCAATACGCT 3354
Db 3277 CGCACAGGCTGCTCTCCCGCGTTTCCAGCGGAGCTGTTCATGATTCCTCAGTATGAT 3336
Qy 3355 ACCTGACGCTCAACATGAGCAGCAAGCGGTGAGAGCTTCACTTTTACGCTGGAAT 3414
Db 3337 ACCTGACGCTTGAACAGAGATCAAGCGGTGAGAGCTTCACTTTTACGCTGGAAT 3396
Qy 3415 ATTTCCCTTCTCAGATGCTGAGAAAGCGGCAACACTTTTACCTTCACTTACCTTGAAG 3474
Db 3397 ACTTCCCTTCTCAGATGCTGAGAAAGCTGAGAAATTAATTTCAATTAAGCTTACCTTGAAG 3456
Qy 3475 AAGTGCCTTTCAACAGACACTACGCGCAGACGCAAGCTTGAAGCGGCTGATGATCTCTC 3534
Db 3457 AAGTGCCTTTCAACAGACACTACGCTCAGCAGAGGTTGATGCTTGAATGATCTCTC 3516
Qy 3535 TCATGACCAATATACCTGATTACTGGAAGAGAACTA---AATAGTCCGGAAGGCGCC 3591
Db 3517 TTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3576
Qy 3592 AAAACAAAGACTTGTGTTTGAAGCGTGGGTCTCAGCTGGCAATGCTGTTGACCCMAA 3651
Db 3577 ACCAATCAAGGCTGCTTTTGAAGCGGCTGAGCTGAGCTGATGCTTGTGACGCGCAGA 3636
```

QY	3652	ACGCGCTACCTGGGACCCCTGTTATGCGGACGAGCGCGTTCTTAAACAAACAAACAGACACA	3711
Db	3637	ATTGGCTACCTGGGCGCCCTGGCTACCGGCAACAGACATTTCAGAGCTGCTAACGACACACA	3696
QY	3712	ACAACAGCAATTTTACTGTGACCTGCTGCTTCAAAATTAACCTCAATGGCGGTGAATCCA	3771
Db	3697	ACAACAGTAATCTTCTCTTGGACAGCGGCGACAAATATCATCTCAATGGCGGCACTCGC	3756
QY	3772	TCATCAACCTTGGCACCTGCTATGGCTTCACACAAAGACGACGAAAGTCTTCCCA	3831
Db	3757	TGGGAATCCAGGACCAAGCTATGGCCAGTCCAAGGACGATGAAGAAAATTTTCCCTTA	3816
QY	3832	TGACCGGCTCATATTTTGGAAAAAGAGCGCGGAGCTTCAACACTGATCTTGACA	3891
Db	3817	TGCACGGCAATCTTAATATTTGGCAAAAGAGGAGCAACGGCAAGTAAGCAGAAATTAGATA	3876
QY	3892	ATGTCATGATTTACAGACGAGAGGAAATTAAGCCACTAACCTCTGTGGCCACGAAAGAT	3951
Db	3877	ATGTAAATGATTAACCGATGAGAGAAAGATTCCTAACCAACCAATCCGTGGCAACAGCAGT	3936
QY	3952	TTGGGACCGTGGGCAATTAATTCAGAGCAGACAGACGACCTCGCACCGGAGATGTGC	4011
Db	3937	ATGGAATCTGTGGCAATTAATCTTCAGAGCTCAAAATCAGCTCCACGACTAGACTGTCA	3996
QY	4012	ATGCTATGGAGCAATTACTTGGCATGTGTGGCAAGATGAGACGTGTACTCTGACGGCTC	4071
Db	3997	ATGATACAGGGGGCTTACCTGGCAATGTGTGGCAAGATCGGAGGTATCTTCAAGAAC	4056
QY	4072	CCATTTGGGCGCAAAATTCCTCACACAGATGACACTTTACCGCTCCTTATGGGCG	4131
Db	4057	CTATCTGGGCAAAAGATTTCTCACAGGAGGACACTTTTCACTCTCTCTGTATGGGAG	4116
QY	4132	GCTTTGCACTCAAGAACCCGCGCTCTCATGATCTCATCAAAACACGCTGTTCCTGCA	4191
Db	4117	GCTTTGCACTCAAAACATCGCGCTCTCAATCATGATCAAAATATCTCCGGTACCGGCAA	4176
QY	4192	ATCTCTCGGCGGAGTTTCACTACAAAGTTTGTCTTCACTACCCAAATATCCCAACAG	4251
Db	4177	ATCTCTCCGACCACTTTCAGCCCGGCAAGTTGTCTCAATTATACCTCAGTACTCCAGT	4236
QY	4252	GACAAGTAGTGTGGAATTTGAATGGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATC	4311
Db	4237	GACAGTAGCGGTGAATTTGAGTGGAGCTACAGAAAGAAACAGCAAGGTTGGAATC	4296
QY	4312	CCGAAGTCAGTACACATCCAAATTATGCAAAATCTGCCAAGCTGATTTTACTGTGACA	4371
Db	4297	CAGAGATTTCAGTACACTTCCAACTACACAAAGTCTGTAACTGTGACTTTACTGTAGACA	4356
QY	4372	ACAATGGACTTTATATGAGGCTGGCGCCATTTGGACCGGTAACCTTACCCGTCGCCGT	4431
Db	4357	CTAATGGGTATATATGTAACCTGGCCCTAATTTGAAACCCGGTATCTCACAGAAACTTGT	4416
QY	4432	AATTACGTGTTAATCAATTAACCGGTTGATTCGTTTCAATTGAATTTGG-TCTCTGTCTC	4490
Db	4417	AA-TCTGTGTTAATCAATTAACCGGTTAATTCGTTTCAATTGAATTTGGCTCTGTGACA	4475
QY	4491	CTTCTTAATCTTATC-GGTTACATAGGTATAGTTACACTTAATCTG-----CTTGGT	4533
Db	4476	CTTCTTAATCTTATCTTGTGTTTCCATGGCTACTGCGTAGATPAACAGCGGCGCTGGCGGCTT	4535
QY	4544	GCGGCTCGCGATTAAGAAC-----TTACGTATCGGGTTACCCCTAGATGATGG	4591
Db	4536	GCGCTTGGGGGTTTACACTGTGCTGTATATTAATTAATCTGGCCATTACTCTAAGATGAG	4599
QY	4592	AGTTGCCCACTTCCTCTCTGCGGCGCTCGCTGCTGTGTGGGACCTGCGGACCAAGGTCC	4651
Db	4586	AGTTGCCCACTCCCTCTTAATGCGCACTCGCTGCTGTGTGGGCGCGGAGTGTGCAAGCAG	4655
QY	4652	GGAGACGGCAAGGCTGTGCTGCGCGGCGCCACGAGGAGGACGACGCGCACAGAGGAGAG	4711
Db	4636	TTCGTTGTGCGACTTTTGTGTGCGCAGGCGCCACACGAGAGAGTGTGTGCTATAGGGAG	4715

[illegible]

Db 583 ACTTCCACATGACGTGCTCTGTGAAACACCGGGGTGAAATCCATGTTTGGGACGTT 642
Qy TCCGTAGATCAGATTAGGGACAAAGCTGGTGTGAGACCATCTACCCGGGATGGAGCCGAC 716
Db 643 TCCGTAGATCAGATTAGGGACAAAGCTGGTGTGAGACCATCTACCCGGGATGGAGCCGAC 702
Qy TGGCCAACTGGTGTGCGGTGACAAAGACGGAATGGCGCGGAGGAGGAAACAAGTTGG 776
Db 703 TGGCCAACTGGTGTGCGGTGACAAAGACGGAATGGCGCGGAGGAGGAAACAAGTTGG 762
Qy TGAACAGTGTATATCCCACTACTCTGTGCCAAGACTCAGCCCGAGCTGACGTGG 836
Db 777 TGAACAGTGTATATCCCACTACTCTGTGCCAAGACTCAGCCCGAGCTGACGTGG 822
Qy TGAATGTGTCTATCCCACTACTCTGTGCCAAGACTCAGCCCGAGCTGACGTGG 896
Db 823 CGTGAATGTGTCTATGGAACAGATTAAAGCGCTGTGTAATCTCAGCAGCGTAAACGCT 882
Qy TCGTGGCGAGCACTGACCCACGTGACGACCAAGCCAGAGCAGAAACAAGAGATCTGA 956
Db 883 TCGTGGCGAGCACTGACCCACGTGACGACCAAGCCAGAGCAGAAACAAGAGATCTGA 942
Qy ACCCCAACTTGTGACGCGCTGTGTATCCGGTCAAAAACCTCCGCGCTCAATGAGCTGG 1016
Db 943 ATCCCAATCTGATGCGCGGTGATCAGATCAAAAACCTCAGCGGTACATGAGCTGG 1002
Qy TCGGATGTGTGAGACCGGGGATCACTCCGGAAGCAGTGAATCCAGAGACCAAG 1076
Db 1003 TCGGATGTGTGAGACCGGGGATCACTCCGGAAGCAGTGAATCCAGAGACCAAG 1062
Qy CCGTGTATCTCTCTCAAGCGCGCTTCAAACTGCGGCTCCAGATCAAGCGCTGG 1136
Db 1063 CCGTGTATCTCTCTCAAGCGCGCTTCAAACTGCGGCTCCAGATCAAGCGCTGG 1122
Qy ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCCGACTACTGTTGAGCCCG 1196
Db 1123 ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCCGACTACTGTTGAGCCCG 1182
Qy CTCGCGCGGCAATTAACCAACCGCATCTACCGCTCTGAGACTGAAGCGCTAG 1256
Db 1183 AGCCCGTGGAGACATTTCCAGCATGGAATTTAAATTTTGAACCTAAACGAGTACG 1242
Qy AACCTGCTACGCGCGCT 1316
Db 1243 ATCCCAATGAGGCTTCT 1302
Qy ACAACATCTGATGTTGGGCGCGGACCAACGCGGCAACCAATGCGGAGGCTAG 1376
Db 1303 ACAACATCTGATGTTGGGCGCGGACCAACGCGGCAACCAATGCGGAGGCTAG 1362
Qy CCGACGCGGCT 1436
Db 1363 CCGACGCGGCT 1422
Qy ATTGCGTGAACAAGATGATCTGTGAGAGAGAGGCAAGATGACGGCAAGGCTGG 1496
Db 1423 ACTGTGTGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1482
Qy AGTCCGCAAGGCT 1556
Db 1483 AGTCCGCAAGGCT 1542
Qy CCGCGCAAGTGAACCCGCT 1616
Db 1543 CCGCGCAAGTGAACCCGCT 1602
Qy TTGAACGGAACAGACCACT 1676
Db 1603 TTGAACGGAACAGACCACT 1662
Qy TTGAATCTACCCCGCT 1736
Db 1663 TTGAATCTACCCCGCT 1722

Qy AGTTCTTCGCTGGGCGCAGGATCAGTGAACGAGGTGGCGATGATTCTACGTGAGA 1796
Db 1723 ACTTTTCCGCTGGGCGCAGGATCAGTGAACGAGGTGGCGATGATTCTACGTGAGA 1782
Qy AGGTTGAGCCCAACAAAGACCCCGCCGATGACCGGATTAAGCCGAGCCGAGCGGG 1856
Db 1783 AGGTTGAGCCCAACAAAGACCCCGCCGATGACCGGATTAAGCCGAGCCGAGCGGG 1842
Qy CCGTCCCT 1916
Db 1843 TGGCGACATGATGCGGACCGCATGACGTCAGACCGGGA--AGCTTCGATCACTACG 1899
Qy CCGACAGGTACCAAAACAATGTTCTGTCAGCGCGGACGATCTCTGATGCTGTTCCCT 1976
Db 1900 CAGACAGGTACCAAAACAATGTTCTGTCAGCGCGGACGATGATGATGCTGTTCCCT 1959
Qy GCAAGACATGCGAGAGATGATCAGAAATTTCAACTTTGCTTCAAGCAGGAGCAGAG 2036
Db 1960 GCAAGACATGCGAGAGATGATCAGAAATTTCAACTTTGCTTCAAGCAGGAGCAGAG 2019
Qy ACTGTTCAAGTGTCTTCCCGCGCTGTCAAGATCTCAACCGGT--GTCAGAAAGAGA 2093
Db 2020 ACTGTTCAAGTGTCTTCC--GTCGTCAGAAATCTCAACCCGTTCTGTCTCAAAAAGG 2076
Qy CGATGCGAAACCTGTGCGATCATCATCTGCTGGGGGGGCTCCCGAGATTGCTTGGT 2153
Db 2077 CGATGCGAAACCTGTGCGATCATCATCTGCTGGGGGGGCTCCCGAGATTGCTTGGT 2133
Qy CGGCTGCGATCTGTCTCAAGCTGAGCCTGAGTACGTGTGTTCTGACAAATAATGACTT 2213
Db 2134 CTGCTGCGATCTGTCTCAAGCTGAGCCTGAGTACGTGTGTTCTGACAAATAATGACTT 2193
Qy AAACCGATGTGCTGCTCGATGTTATCTTCCAGATTGCTGAGAGCAACTCTCTGAG 2273
Db 2214 AAACCGATGTGCTGCTCGATGTTATCTTCCAGATTGCTGAGAGCAACTCTCTGAG 2253
Qy AATTCAGATGTGCTGCTCGATGTTATCTTCCAGATTGCTGAGAGCAACTCTCTGAG 2274
Db 2274 AATTCAGATGTGCTGCTCGATGTTATCTTCCAGATTGCTGAGAGCAACTCTCTGAG 2253
Qy GGCATTCGAGAGTGTGGGCTTGAACCTGAGACCCGAGCCGAAAGCCAAAGCCAGCAA 2333
Db 2254 GGCATTCGAGAGTGTGGGCTTGAACCTGAGACCCGAGCCGAGCCGAAAGCCGAGCAA 2313
Qy AAGCAGACGACGCGCGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2393
Db 2314 AAGCAGACGACGCGCGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2373
Qy GGACTGACAGGGGAGCCCTCTCAACGCGGAGACGACGCGGCTCTCAAGCAGCAAG 2453
Db 2374 GGACTGACAGGGGAGCCCTCTCAACGCGGAGACGACGCGGCTCTCAAGCAGCAAG 2433
Qy GCTTACGACGACGCTCAAAAGCGGCTGCAATCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
Db 2434 GCTTACGACGACGCTCAAAAGCGGCTGCAATCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2493
Qy GCCGAGTTTCAGAGCGTGTGCAAGAGATACGCTTTTGGGGGCAACTCGGGCGAGCA 2573
Db 2494 GCCGAGTTTCAGAGCGTGTGCAAGAGATACGCTTTTGGGGGCAACTCGGGCGAGCA 2553
Qy GTTTCAGGCGCAAGAGCGGCTTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2633
Db 2574 GTTTCAGGCGCAAGAGCGGCTTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2613
Qy GTTTCAGGCGCAAGAGCGGCTTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2693
Db 2614 GTTTCAGGCGCAAGAGCGGCTTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2673
Qy GGCATGCGGCAAGACGAGCCGCTTAAGAAAGACCTCAATTTTGGTCACTGAGCTGCG 2753
Db 2674 GGCATGCGGCAAGACGAGCCGCTTAAGAAAGACCTCAATTTTGGTCACTGAGCTGCG 2733
Qy GACTCAGATCACT 2813
Db 2734 GACTCAGATCACT 2793

QY 2814 GTGGACCTACTAATGCTTCAAGCGGTGGCGCAACCAATGAGACAAATACGAAGGC 2873
 DB 2794 CTGGGAAGTATATGATGGCTTACAGGCAAGTGGCGCAACCAATGAGAGGCG 2853
 QY 2874 GCCGACGAGTGGGTATGCTTCAAGAAATTGGCAATTCGATTCACATGCTGGCGAC 2933
 DB 2854 GCCGACGAGTGGGTATGCTTCAAGAAATTGGCAATTCGATTCACATGATGGCGAC 2933
 QY 2934 AGATCAATCACCACGAGACCGGACCTGGGCTTGGCCACTTCAATTAACCACTCTAC 2993
 DB 2914 AGATCAATCACCACGAGACCGGACCTGGGCTTGGCCACTTCAATTAACCACTCTAC 2993
 QY 2994 AAGCAATCTCAGTGTCTTCAAGCGGCGCAGCAACCACTACTTGGGCTAGAC 3053
 DB 2974 AAGCAATCTCAGTGTCTTCAAGCGGCGCAGCAACCACTACTTGGGCTAGAC 3053
 QY 3054 ACCCTTGGGGGTATTTTGAATTTCAAGATTTCACTGCTTCACTTTTCACTGCTG 3113
 DB 3031 ACCCTTGGGGGTATTTTGAATTTCAAGATTTCACTGCTTCACTTTTCACTGCTG 3090
 QY 3114 CAGGACCTACCAACCAATTTGGGGATTTGGGCGCAAGACCTCAACTTCAACTCTTC 3173
 DB 3091 CAAAGCTCATCAACCAACTGGGGATTTGGGCGCAAGACCTCAACTTCAACTCTTC 3150
 QY 3174 AACATCCAGTCAAGAGGTCAAGCAAGATGAGCTTCAACACATCGCTAATTAACCTT 3233
 DB 3151 AACATCCAGTCAAGAGGTCAAGCAAGATGAGCTTCAACACATCGCTAATTAACCTT 3210
 QY 3234 ACCGACGCTTCAAGTCTTCTCGGACCTGGAATCCAGCTTCCGTAAGTCTTGGCTCT 3293
 DB 3211 ACCGACGCTTCAAGTCTTCTCGGACCTGGAATCCAGCTTCCGTAAGTCTTGGCTCT 3270
 QY 3294 GCGACGAGGCTGCTCTCCCTCCGTTCCGCGGACGCTTCAATGATTCGCAATTAACGCG 3353
 DB 3271 GCGACGAGGCTGCTCTCCCTCCGTTCCGCGGACGCTTCAATGATTCGCAATTAACGCG 3330
 QY 3354 TACCTGAGCTCAACATGAGGAGCCAGCCGCTGGACCTTCACTTCTTCTGCTGAA 3413
 DB 3331 TACCTGAGCTCAACATGAGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3390
 QY 3414 TATTTCTCTTCAAGTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3473
 DB 3391 TATTTCTCTTCAAGTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3450
 QY 3474 GAGGCTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3533
 DB 3451 GAGGCTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3510
 QY 3534 CTCACTGACCAATACCTGATTAACCTGAAAGAACTCAAAATCAAGTCCGAAAGTCCCA 3593
 DB 3511 CTCACTGACCAATACCTGATTAACCTGAAAGAACTCAAAATCAAGTCCGAAAGTCCCA 3570
 QY 3594 AACAGGAGCTTGTGTTAGGCGGTGCTCAGCTGCAATGTCTGTTCAAGCCCAAAAC 3653
 DB 3571 CAGTCAAGGCTTGTGTTAGGCGGTGCTCAGCTGCAATGTCTGTTCAAGCCCAAAAC 3630
 QY 3654 TGGCTACCTGAGACCTGTTATCGGACGAGCGGCTTTTAAACCAAAACAGACCAAC 3713
 DB 3631 TGGCTACCTGAGACCTGTTATCGGACGAGCGGCTTTTAAACCAAAACAGACCAAC 3690
 QY 3714 AACAGCAATTTTACCTGAGCTGTGCTTCAAAATTAACCTCAATGGCGGTAAATCCATC 3773
 DB 3691 AACAGCAATTTTACCTGAGCTGTGCTTCAAAATTAACCTCAATGGCGGTAAATCCATC 3750
 QY 3774 ATCAACCTGAGCACTGCTATGCTTCAACCAAAAGCAAGCAAGTCTTTTCCATG 3833
 DB 3751 GTGATTCGAGCGCGGCGCATGCAAGCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3810
 QY 3834 AGCGGTGCTGATTTTGAAGAGAGCGCGGAGCTTCAACCACTGCAATGAGCAAT 3893
 DB 3811 AGCGGTGCTGATTTTGAAGAGAGCGCGGAGCTTCAACCACTGCAATGAGCAAT 3870
 QY 3894 GTCAATGATTAAGATTT 3953

DB 3871 GTCAATGATTAAGAT 3930
 QY 3954 GGAACCTGAGCAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4013
 DB 3931 GGTCTGTATTAACCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3990
 QY 4014 GGTATGAGCAATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4073
 DB 3991 ACACAGGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4050
 QY 4074 ATTTGGGCAAAATTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4133
 DB 4051 ATCTGGGCAAAATTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4110
 QY 4134 TTTGAGTCAAGAT 4193
 DB 4111 TTGGAGTCAAGAT 4170
 QY 4194 CTTCCGCGGAGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4253
 DB 4171 CTTCCGAGCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4230
 QY 4254 CAAGTGTGTAATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4313
 DB 4231 CAGGTCAAGTGTGTAATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4290
 QY 4314 GAAGTGTGTAATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4373
 DB 4291 GAATGTGTAATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4350
 QY 4374 AATGACTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4433
 DB 4351 AATGAGTGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4410
 QY 4434 TTACGTGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4493
 DB 4411 TTGCTGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4470
 QY 4494 CTTATCTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4553
 DB 4471 TCTTTCTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4530
 QY 4554 ATAAAGACTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4613
 DB 4531 TACAGAT 4574
 QY 4614 CGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 4673
 DB 4575 CGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 4634
 QY 4674 GCGGCGGAT 4718
 DB 4635 GCGGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4679

RESULT 6
 AF043303 4679 bp DNA linear VRL 24-FEB-1998
 LOCUS AF043303
 DEFINITION Adeno-associated virus 2, complete genome.
 ACCESSION AF043303
 VERSION AF043303.1 GI:2906016
 KEYWORDS
 SOURCE
 ORGANISM
 adeno-associated virus 2.
 Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 REFERENCE
 1 (bases 1 to 4679)
 Rufing, M., Heid, H., and Kleinschmidt, J. A.
 Mutations in the carboxy terminus of adeno-associated virus 2
 capsid proteins affect viral infectivity: lack of an RGD
 integrin-binding motif
 JOURNAL U. Gen. Virol. 75 (Pt 12), 3385-3392 (1994)
 MEDLINE 95088582

PUBMED 7996133
REFERENCE 2 (bases 1 to 4679)
AUTHORS Barnes, K.I., Bohenzky, R.A., Cassinotti, P., Colvin, D., Donahue, B.A.,
Dill, T., Horer, M., Kleinschmidt, J.A., Ruffing, M., Snyder, R.O.,
Tratschin, J.-D. and Weitz, M.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1998) Cell Genesys Inc., 342 Lakeside Dr., Foster
City, CA 94404, USA

FEATURES
source
1. .4679
/organism="adeno-associated virus 2"
/db_xref="taxon:10804"
/note="changes relative to the original sequence, GenBank
Accession Number J01901, have been detected and verified
by several different laboratories"
1. .145
/note="inverted terminal repeat"
/rpt_type=inverted
42. .83
/note="flip oriented DNA"
287. .4451
/codon_start=1
/product="Rep 68 protein"
/protein_id="AAC03774.1"
/db_xref="GI:2906017"
/translation="MGFVEIVIKVPSDLDEHLPGISDFVNVAAEKEMELPPDSMD
LNIHQAPLTVAKLQRPDLETRRYSKAEALFPQPEKESYFHKHVLVETTKYS
MVGRLFSQIREKLIQRIYRGIPTLPNPAVTKTRNGAGGKRVDECVIENYLK
TOPELQAMWNTMEOYLSACNLTERKLVNQHLYHSQTEOKNENPNPSDAPVRS
KTSARVYELVGMVDKGTISEKQIODEQASYSFNAASRSQIKALDNACKTMSL
TKTAPYLVGOQPEVEDISSNRKYKILENGYDQVYASVFLGNATKKFGKRNITWLG
ILGSKRVNDQCKSSAQIDPTPIVITSNTMCAVIDGNTTFEHOQPLDDRMKEFL
TRLLDHPGKVTQOEVDFFRMAKDHVEVEHEFPYKGGAKKRPAASDADISEPRV
RESVAQSTSDAASINAYRYONKCSRHVGMMLMFPCCQCRNANONSICTHGOK
DCLCEFPVESQSPVSVAKKAYOKLCYIHHIMGVDPDCTACDLVNVLDLCITEQ"
370
/codon_start=1
/product="Rep 78 protein"
/protein_id="AAC03775.1"
/db_xref="GI:2906018"
/translation="MGFVEIVIKVPSDLDEHLPGISDFVNVAAEKEMELPPDSMD
LNIHQAPLTVAKLQRPDLETRRYSKAEALFPQPEKESYFHKHVLVETTKYS
MVGRLFSQIREKLIQRIYRGIPTLPNPAVTKTRNGAGGKRVDECVIENYLK
TOPELQAMWNTMEOYLSACNLTERKLVNQHLYHSQTEOKNENPNPSDAPVRS
KTSARVYELVGMVDKGTISEKQIODEQASYSFNAASRSQIKALDNACKTMSL
TKTAPYLVGOQPEVEDISSNRKYKILENGYDQVYASVFLGNATKKFGKRNITWLG
ILGSKRVNDQCKSSAQIDPTPIVITSNTMCAVIDGNTTFEHOQPLDDRMKEFL
TRLLDHPGKVTQOEVDFFRMAKDHVEVEHEFPYKGGAKKRPAASDADISEPRV
RESVAQSTSDAASINAYRYONKCSRHVGMMLMFPCCQCRNANONSICTHGOK
DCLCEFPVESQSPVSVAKKAYOKLCYIHHIMGVDPDCTACDLVNVLDLCITEQ"
370
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="g"
878. .4451
/codon_start=1
/product="Rep 40 protein"
/protein_id="AAC03776.1"
/db_xref="GI:2906019"
/translation="MELVGMVLVDKGTISEKQIODEQASYSFNAASRSQIKALD
NCKTMSLTKTAPDYLVGOQPEVEDISSNRKYKILENGYDQVYASVFLGNATKKFGK
RNITWLGKATITGKTNIAEALIAHTVPGCVNNTNENPPNDQVDMKVIWMEGKMTA
KVESAKAILDGSKRVNDQCKSSAQIDPTPIVITSNTMCAVIDGNTTFEHOQPLD
DRMKEFLTRLLDHPGKVTQOEVDFFRMAKDHVEVEHEFPYKGGAKKRPAASDA
DISEPRVRESVAQSTSDAASINAYRYONKCSRHVGMMLMFPCCQCRNANONSICTHGOK
993. .2186
/codon_start=1
/product="Rep 52 protein"
/protein_id="AAC03777.1"
/db_xref="GI:2906020"

variation

precursor_RNA
Intron
CDS
1853. .4451
1907. .2227
2203. .4410
/codon_start=1
/product="major coat protein VP1"
/protein_id="AAC03780.1"
/db_xref="GI:2906023"
/translation="MADGILPDWLEDTLSEGIQWKKLKPGEPPKPAERHKDSRG
LVLPYKYLQGFENGLDGEPEVNEADAALAEHDAVRODQSGNPLKTNHAAEFOE
RLKEDTSFGGILGRAVFOAKRVLELIGVEEVPKTAQPKRVERSPVPEDSGSGTQ
KAGQOPARKRLNGOTGDADSVDPDQLOPPAPASGLGNTMAQASGAMADNNGA
DGVNSSGNMHCSTMMGDRVITTSRTMALPTVNNHLYKQISSQASADNHPGYS
TPWGYDPNRHCHFSRPRDMORLINNMGFRPKRLNPLNIOYKEXTONDGTTTAN
NLTSIVQFTDSEYQLPYVDSAHQCLPEPPADVWVPOYGLITLNGSOAVGRSSF
YCLEYPSQMLRTGNNTFSYTFEDVFFHSYVASHQSLRLNMLIDYLYLSRNT
PSCGTTQSLQFQAGASDIRQSRMLPGPCYQOQSVKTSADNNSSEYSGATKY
HLNGRDSLVPNGPAMASHKODEEKFPPQSGVILFGKQSEKTVNDEKWIITDEEIR
TNPNVATQYGSVSTNLQRGNQOATADVNTQGVLRGMWQODRDVYLQIMAKIBHT
DGHFHSPLMGSGGLGHPPOILLIKATPPANSTTFSAKRPASFTIQYSTQSVSEI
EMELQKENSRRNMBEIQYTSNKSANVDPVDTNVSSEPRIGTRYLTRNL"
2429
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="ta"
2614. .4410
/codon_start=1
/transl_except=(pos:2614. .2616, aa:Met)
/product="major coat protein VP2"
/protein_id="AAC03778.1"
/db_xref="GI:2906021"
/translation="NAPGKKRPVHSPVPEPDSGSGTKAGQOPARKRLNFGOTGDAD
VDPDQLOPPAPASGLGNTMAQASGAMADNNGAAGVNSGNMHCSTMMGDRV
ITTSRTMALPTVNNHLYKQISSQASADNHPGYSSTMGVDFNRHCHFSRDMQ
RLINNMGFRPKRLNPLNIOYKEXTONDGTTTANNTSVQVFTDSEYQLPYV
SHQGLPEPPADVWVPOYGLITLNGSOAVGRSSFYLEYPSQMLRTGNNTFSY
TFEDVFFHSYVASHQSLRLNMLIDYLYLSRNTPSCGTTQSLQFQAGASDIR
QSRMLPGPCYQOQSVKTSADNNSSEYSGATKYHLNGRDSLVPNGPAMASHKOD
EEKFPPQSGVILFGKQSEKTVNDEKWIITDEEIRTNPNVATQYGSVSTNLQRGN
QOATADVNTQGVLRGMWQODRDVYLQIMAKIBHTDGHFHSPLMGSGGLGHPPO
ILLIKATPPANSTTFSAKRPASFTIQYSTQSVSEIEMELQKENSRRNMBEIQYTSN
KSNVAVDPVDTNVSSEPRIGTRYLTRNL"
2809. .4410
/codon_start=1
/product="major coat protein VP3"
/protein_id="AAC03779.1"
/db_xref="GI:2906022"
/translation="MATSGAPPADNNEGADGVNSGNMHCSTMMGDRVITTSRT
MALPTVNNHLYKQISSQASADNHPGYSSTMGVDFNRHCHFSRPRDMORLINN
MGFRPKRLNPLNIOYKEXTONDGTTTANNTSVQVFTDSEYQLPYVDSAHQCL
PEPPADVWVPOYGLITLNGSOAVGRSSFYLEYPSQMLRTGNNTFSYTFEDV
HSSYVASHQSLRLNMLIDYLYLSRNTPSCGTTQSLQFQAGASDIRQSRML
PGPCYQOQSVKTSADNNSSEYSGATKYHLNGRDSLVPNGPAMASHKODEEKF
PPQSGVILFGKQSEKTVNDEKWIITDEEIRTNPNVATQYGSVSTNLQRGNQOAT
ADVNTQGVLRGMWQODRDVYLQIMAKIBHTDGHFHSPLMGSGGLGHPPOILLIK
ATPPANSTTFSAKRPASFTIQYSTQSVSEIEMELQKENSRRNMBEIQYTSNKSANV
DPVDTNVSSEPRIGTRYLTRNL"
2877
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="c"
3759. .3765
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="g"
3859

variation

variation

variation

/note="compared to sequence of GenBank Accession Number
J01901"
/replace="a"
3898
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="n"
3900..3902
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="gaac"
4232..4233
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="acg"
4329..4330
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="tcg"
4336
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="c"
4347
/note="compared to sequence of GenBank Accession Number
J01901"
/replace="t"
4535..4679
/note="inverted terminal repeat"
/rpt_type=inverted
4597..4638
/note="flop oriented DNA"
misc_feature
1198 a 1262 c 1255 g 964 t
BASE COUNT
ORIGIN
Query Match 65.9%; Score 3107.8; DB 14; Length 4679;
Best Local Similarity 80.2%; Pred. No. 0;
Matches 3790; Conservative 0; Mismatches 882; Indels 53; Gaps 10;

403 ACTGGGTGGCCGAGAAAGAAATGGAGATTGCGGCCGCAATTCGACATGATCTGAATCTGA 462
477 TTGAGCAGGCAACCCCTGACCGTGGCCGAGAACTGCAAGCGGACTTCCTGTGTCATATGC 536
463 TTGAGCAGGCAACCCCTGACCGTGGCCGAGAAAGTGTGACGCCGACTTCTGACGGAATGCG 522
537 GCCCGGTGAATAGGCCCGGAGAGCCCTCTCTTTGTTCAGTTCGAAAGGCGAGCT 596
523 GCGGTGTGAATAGGCCCGGAGAGCCCTCTCTTTGTTCAGTTCGAAAGGCGAGAGCT 582
597 ACTTCACTTCATATTCTGTGTGAGAACACGCGGGGTCAATCATGATGCTGGCCGCT 656
583 ACTTCACTGACAGTGTGTGTGAGAACACGCGGGGTCAATCATGATGCTGGGACGTT 642
657 TCCTGATGATTTAGGACAACTGTGTGAGAACATCTACCCCGGGATGAGCCGACCC 716
643 TCCTGATGATTTAGGACAACTGTGTGAGAAATTTACCCCGGGATGAGCCGACCTT 702
717 TGCCCAACTGTTTCGCGGTGACCAAGCGGTATGCGCGCGGAGGGGGGAGACAAAGTGG 776
703 TGCCCAACTGTTTCGCGGTGACCAAGACCAAGATGCGCGCGGAGGGGAGACAAAGTGG 762
777 TGAAGAGTGTACATCCCACTACTCTGTGCCAAGACTGACCCGAGCTGCAATGG 836
763 TGAATGATGTCTACATCCCACTACTCTGTGCCAAGACTGACCCGAGCTGCAATGG 822
837 CGTGAATACATGAGAGATATATAGGCGCTGTGAACTGGCCGCGGACGGAACCGGC 896
823 CGTGAATATATATAGAAACATATTTAAGCGCTGTGAACTTCACGAGACGTAAACGT 882
897 TCGTGGCGCAGCACTGACCCGACGTCAGCAGACCCAGAGAGCAAAACAGAGAAATCTGA 956
883 TGTGTGGCGCAGCATCTGACCGCAGCTGTGCGAGACGAGAGAGCAAGAAAGATATAGA 942
957 ACCCAATTCTGACGCGCTGTATCCGATCCGATCAAAAACTTCGCGCGCTACATGAGCTGG 1016
943 ATCCCAATTCTGATGCGCGGTATGATGATGATGATGATGATGATGATGATGATGATG 1002
1017 TCGGTGGCTGTGTGACCGCGGATCATCTCCGAGAAAGATGATGATGATGATGATGATG 1076
1003 TCGGTGGCTGTGTGACCGCGGATCATCTCCGAGAAAGATGATGATGATGATGATGATG 1062
1077 CTTGTATCATCTCTTCAACGCGCTTCAACTCGCGGTTCAGATCAAGCGCTCTGG 1136
1063 CTTGTATCATCTCTTCAACGCGGCTTCAACTCGCGGTTCAGATCAAGCGCTCTGG 1122
1137 ACAATGCGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196
1123 ACAATGCGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1182
1197 CTCGCGCGGAGCATTTAAACCAACGCGATCTACCGCATCTGAGAGTGAACGGCTAGC 1256
1183 AGCCCGTGAAGACATTTCCAGCAATGATGATTTATTAATTTTGAATTAACGGGTAGC 1242
1257 AACCTGCTACGCGGCTCCGCTTTCTCGGCTGGGCCGAGAAAGGTTGGAGAGCGCA 1316
1243 ATCCCAATATGAGGCTCCGCTTTCTCGGAGTGGGCAAGAAAGTGGGCAAGAGGA 1302
1317 ACAACATCTGCTGTTTGGCGCGGCAACGCGGCAAGACCAATCTGCGGAAAGCCATCG 1376
1303 ACAACATCTGCTGTTTGGCGCGGCAACGCGGCAAGACCAATCTGCGGAAAGCCATCG 1362
1377 CCCAGCGCGTGCCTTCAAGGCTGAGTCAATGAGCAATGAGAACTTCCCTTCAATG 1436
1363 CCCAGCTGCGCTTCAAGGCTGAGTCAATGAGCAATGAGAACTTCCCTTCAAGC 1422
1437 ATTGCGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1496
1423 ACTGTGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1482
1497 AGTCCGCAAGGCAATCTTCCGCGGAGCAAGGTGCGCTGTGACCAAAATGCAAGTCTGT 1556
1483 AGTCCGCAAGGCAATCTTCCGCGGAGCAAGGTGCGCTGTGACCAAAATGCAAGTCTGT 1542

QY 1557 CCGCCGATCGACCCCGCGTATCGTCACTTCAACCAACATGTGCGGTGA 1616
DB 1543 CGGCCGAGATGAGACCCGACTCCCGTATCGTCACTTCAACCAACATGTGCGGTGA 1602
QY 1617 TTGACGGGAAAGACCACTTTCGAGACGAGCGGTTCGAGACCGATGTTCAAT 1676
DB 1603 TTGACGGGAACTCAACGACCTTTCGACCAAGGCGGTTCGAGACCGATGTTCAAT 1662
QY 1677 TTGACCTACCCCGCGTTCGAGACGATCTTTCGAGGAGTGAAGAGAGAGTCAAG 1736
DB 1663 TTGAACTACCCCGCGTTCGATCTGATCTTTCGAGGAGTGAAGAGAGAGTCAAG 1722
QY 1737 AGTTCTTCGCTGGGCGCGATCACTGACCGAGTGGCGCATGATGTTCAAGTGA 1796
DB 1723 ACTTTTCGCGTGGGCAAGATCACTGATGAGTGGAGCATGATGTTCAAGTGA 1782
QY 1797 AGGGTGAAGCAAAAGACCCGCGCGATGAAGCGATTAAGCGAGCCCAAGCGG 1856
DB 1783 AGGGTGAAGCAAAAGACCCGCGCGATGAAGCGATTAAGCGAGCCCAAGCGG 1842
QY 1857 CCGCCCGTATCGCGGATTCATCGACGTCAGACCGGAGAGAGTCCGCGTGA 1916
DB 1843 TCGCGAGTCAAGTTCGCGACCATCGACGTCAGACCGGAGAGAGTCCGCGTGA 1899
QY 1917 CCGACAGATCAAAACCAATGTTCTCGTACGCGCGGATGCTTCAAGTCTGTTCC 1976
DB 1900 CAGACAGATCAAAACCAATGTTCTCGTACGCGCGGATGCTTCAAGTCTGTTCC 1959
QY 1977 GCAAGACATCGAGAGATGAATCAATTTCAATTTCTTCAAGCGAGCGAG 2036
DB 1960 GCAAGACATCGAGAGATGAATCAATTTCAATTTCTTCAAGCGAGCGAG 2019
QY 2037 ACTGTTCAAGTCTTCCCGCGGTTCAGATTTCAAGCGGT---CGTCAAGAGAG 2093
DB 2020 ACTGTTCAAGTCTTCCCGCGGTTCAGATTTCAAGCGGT---CGTCAAGAGAG 2076
QY 2094 CGTATCGGAACCTCTGCGCATTCATCTGCGGCGGCGGCGCGGATGCTTCT 2153
DB 2077 CGTATCGGAACCTCTGCGCATTCATCTGCGGCGGCGGCGGATGCTTCT 2133
QY 2154 CGGCGTGCATCTGTCAGCGTGAAGCTGATGAATGTTCTGAGCAATTAATGA 2213
DB 2134 CTGCGTGCATCTGTCAGCGTGAAGCTGATGAATGTTCTGAGCAATTAATGA 2193
QY 2214 AAACGAGATGCTGCGGATGTTATCTTCAAGTTCGATGCTGAGCAACTCTGAG 2273
DB 2194 AAATCAAGTTCGCTGCGGATGTTATCTTCAAGTTCGATGCTGAGCAACTCTGAG 2253
QY 2274 GGCATTCGCGAGTGGTGGGACTTGAACCTGAGCCCGAGCCCAAGCGAGCA 2333
DB 2254 GGAATTAAGAGTGGTGGGAGCTCAACCTGAGCCCGAGCCCGAGCGG 2313
QY 2334 AAGCAGAGCAGCGCGCGGTCTGTCCTTCTGCTCAAGTACCTCGAGCCCTTCA 2393
DB 2314 CATTAAGAGCAGCGCGGTCTTGTGCTTCTGAGGATCAAGTACCTCGAGCCCTTCA 2373
QY 2394 GGAATCGAGAGGAGCGCGGTTCAGCGCGGAGCGAGCGCGCTTCGAGCAAG 2453
DB 2374 GGAATCGAGAGGAGCGCGGTTCAGCGCGGAGCGAGCGCGCTTCGAGCAAG 2433
QY 2454 GGCCTACGACGAGCCTCAAGCGGATGATCAATCGGATGCGGATTAAGCGCGG 2513
DB 2434 GGCCTACGACGAGCCTCAAGCGGATGATCAATCGGATGCGGATTAAGCGCGG 2493
QY 2514 GCGGAGTTTCAGAGCGTTCGAGAGAGATGCTTTCGAGGAGCAACTCGGCGAGCA 2573
DB 2494 GCGGAGTTTCAGAGCGCTTAAAGAGATACGCTTTCGAGGAGCAACTCGGCGAGCA 2553
QY 2574 GTCTTCAGAGCGAGAGCGGTTCGAGAGCTTCTGCTGCTGAGAGAGCGCTAG 2633
DB 2554 GTCTTCAGAGCGAGAGAGCGGTTCGAGAGCTTCTGCTGCTGAGAGAGCGCTAG 2613

QY 2634 ACCGCTCTCGAAGAAAGAGCGTGGTGAAGAGTGGCGCAAGAGCGAGACTCTCTCG 2693
DB 2614 ACCGCTCTCGAAGAAAGAGCGTGGTGAAGAGTGGCGCAAGAGCGAGACTCTCTCG 2673
QY 2694 GGCATGCGAGAGCAGGCGAGAGCGCGTAAAGAGACTCAATTTGCTGAGACTGCG 2753
DB 2674 GGAACCGAAGAGCGGCGAGAGCGTGAAGAAAGATTTGATTTGCTGAGACTGGA 2733
QY 2754 GACTCAGATGCTGCGGATTCAGAACTCTCGGAGAACTCCAGCAACCCCGCTGCT 2813
DB 2734 GAGCAGACTCAGTACTTACCCCGCTCTCGAGACCGACCGACCGCGCTCTGCT 2793
QY 2814 GTGGGACTACTACATGCTTCAGGCGGTGGCGACCATGGCAGACATTAAGAGG 2873
DB 2794 CTGGGAACTAATGCTTACGAGTACGAGCATGGGCGACCATGGCAGACATTAAGAGG 2853
QY 2874 GCGGACGAGTGGGATGCTTCAGGAAATGGCATTTGCGATTCAGATGCTGGCGAG 2933
DB 2854 GCGGACGAGTGGGATGCTTCAGGAAATGGCATTTGCGATTCAGATGCTGGCGAG 2913
QY 2934 AGAGTATGACCAACGAGACCCGACCTGGGCTGGCCACCTACATTAAGCACTTAC 2993
DB 2914 AGAGTATGACCAACGAGACCCGACCTGGGCTGGCCACCTACATTAAGCACTTAC 2973
QY 2994 AAGCAATCTCCAGTCTTCAAGCGGCGCAGCAAGCAACCACTTCTGCTACAG 3053
DB 2974 AAGCAATCTCCAGTCTTCAAGCGGCGCAGCAAGCAACCACTTCTGCTACAG 3030
QY 3054 ACCCGTGGGAGTATTTGATTTCAAGATTCAGATTCAGCTTTCACACGTCG 3113
DB 3031 ACCCGTGGGAGTATTTGATTTCAAGATTCAGATTCAGCTTTCACACGTCG 3090
QY 3114 CAGCGATGATCAAGCAAGTGGGATTCGCGCCCAAGACTGAATTCATTAACCT 3173
DB 3091 CAAGATGATCAAGCAAGTGGGATTCGCGCCCAAGACTGAATTCATTAACCT 3150
QY 3174 AACATCAAGTCAAGAGGTCAGAGATGATGCGTCAACCACTCTATTAACCT 3233
DB 3151 AACATCAAGTCAAGAGGTCAGAGATGATGCGTCAACCACTCTATTAACCT 3210
QY 3234 ACCAGCAGTTCAGATCTTTCGAGCTGAGATCAAGCTTCGATCTCTGCTCT 3293
DB 3211 ACCAGCAGTTCAGATCTTTCGAGCTGAGATCAAGCTTCGATCTCTGCTCT 3270
QY 3294 GCGCAGCAGGCTGCTCCGCTCCGCTCCGCGGAGAGTGTATGATTCGGAATAGCG 3353
DB 3271 GCGCAGCAGGAGTCTCCGCTCCGCTCCGCGGAGAGTGTATGATTCGGAATAG 3330
QY 3354 TACCTGAGCTCAACATGAGCAGAGCGGAGGCTTCACTTACCTTACTGCTGAA 3413
DB 3331 TACCTGAGCTCAACATGAGCAGAGCGGAGGCTTCACTTACCTTACTGCTGAA 3390
QY 3414 TATTTCCCTTCAGATGCTGAGAGCGGCAACATTTACCTTACCTTACTGCTGAA 3473
DB 3391 TATTTCCCTTCAGATGCTGAGAGCGGCAACATTTACCTTACCTTACTGCTGAA 3450
QY 3474 GAAATGCTTTCAGAGAGCTTACGCGGACAGCTGAGAGCTTGAAGCGGCTGATGATCT 3533
DB 3451 GAAATGCTTTCAGAGAGCTTACGCGGACAGCTGAGAGCTTGAAGCGGCTGATGATCT 3510
QY 3534 CTGATGAGCAATTAAGCTGATTAAGTGAAGCAAGTCAATTAAGTGAAGTGGCCAA 3593
DB 3511 CTGATGAGCAATTAAGCTGATTAAGTGAAGCAAGTCAATTAAGTGAAGTGGCCAA 3570
QY 3594 AACAGAGCTTGTGTTAGCGGTGGCTTCAAGCTGAGTGTGTTAGCGCCCAAAA 3653
DB 3571 CAGTCAAGCTTCAAGTTTCTCAGGCGGAGGAGGATTCAGGAGCAAGTGTAGAAC 3630
QY 3654 TGGCTACCTGAGACCGCTTATGAGGAGAGCGGCTTTCATTAAGCAAGCAAGCAAC 3713
DB 3631 TGGCTACCTGAGACCGCTTATGAGGAGAGCGGCTTTCATTAAGCAAGCAAGCAAC 3690
QY 3714 AACAGCAATTTACCTGAGCTGCTTCAAAATTAACCTCAATGAGGCTGATTCATC 3773

QY 777 TGAACGAGTCTACATCCCACTACCTCTGCGCAAGACTCAGCCGAGCTGAGTGG 836
DB 763 TGAATGAGTCTACATCCCACTACCTCTGCGCAAGACTCAGCCGAGCTGAGTGG 822
QY 837 CGTGAATCAATGAGAGATATATAGAGCCCTGTTTGAACCTGGCGAGCGCAACGCG 896
DB 823 CGTGAATCAATGAGAGATATATAGAGCCCTGTTTGAACCTGGCGAGCGCAACGCG 882
QY 897 TCGTGGCGGAGCACTGACCCAGTCAAGCCAGAGCCAGAGAGCAAGAGATCTGA 956
DB 883 TCGTGGCGGAGCACTGACCCAGTCAAGCCAGAGCCAGAGAGCAAGAGATCTGA 942
QY 957 ACCCAATTTCTGACGCGCTGTGATCCGCTCAAAAACCTCGCGCGCTACATGAGAGTGG 1016
DB 943 ATCCCAATTTCTGATGCGCGCTGTGATCAAGATCAAAAACCTCAAGCGATACATGAGAGTGG 1002
QY 1017 TCGGAGTGGCTGATGAGACCGGAGGATCACTCCGAGAGAGAGTGGATCCAGAGAGCCAGG 1076
DB 1003 TCGGAGTGGCTGATGAGACCGGAGGATTAATCTCGGAGAGAGAGTGGATCCAGAGAGCCAGG 1062
QY 1077 CCTGATCAATCTCTCTCAACGCGCTTCCAACTGCGCGCTCCAGATCAAGCGCGCTCTGG 1136
DB 1063 CCTCAATACATCTCTCTCAACGCGCTTCCAACTGCGCGCTCCAAATCAAGCGCTGCTTGG 1122
QY 1137 ACAATGCGGCAAGATCATGAGCGCTGACCAATCCGCGCGCGCTACCTGCTGAGGCGCG 1196
DB 1123 ACAATGCGGCAAGATCATGAGCGCTGACCAATCCGCGCGCGCTACCTGCTGAGGCGCG 1182
QY 1197 CTCGCGCGCGGACATTTAAACCAACCGCATCTACCGCATCTGAGAGCTGAACGCGTACG 1256
DB 1183 AGCCCGTGGAGGACATTTCAAGCAATCGATTTATTAATTTTGAACCTAAACGCGGTAAG 1242
QY 1257 AACCTGCTACGCGCGCTCGCTCTTTCTGCGCTGGGCGCCAGAAAGGTTGGGAAAGCGCA 1316
DB 1243 ATCCCAATTAATGAGGCTTCCGCTTTCTGAGATGGGCGCAAGAAAGTTGCGCAAGAGGA 1302
QY 1317 ACACCAATCTGAGCTGTTTGGGCGGCGCACACGCGGCAAGACCAATCGCGGAGAGCCATCG 1376
DB 1303 ACACCAATCTGAGCTGTTTGGGCGGCTGACCACTACCGGAGAGACCAATCGCGGAGAGCCATCG 1362
QY 1377 CCCAGCGCGGCTCTCTCAAGGCTGCGTCAACTGAGCAATGAGAACTTTCCCTTCAATG 1436
DB 1363 CCCAGCGCGGCTCTCTCAAGGCTGCGTCAACTGAGCAATGAGAACTTTCCCTTCAAG 1422
QY 1437 ATTGCGTCAAGATGATGATCTGATGGAGAGAGGCAAGATGACGCGCAAGTGTGG 1496
DB 1423 ACTGCTGCAAGATGATGATCTGATGGAGAGAGGCAAGATGACGCGCAAGTGTGG 1482
QY 1497 AGTCCGCAAGGCGCATCTGCGGCGGCGCAAGAGTGGCGTGGACCAAAAGTGAAGTGGT 1556
DB 1483 AGTCCGCAAGGCGCATCTGCGGCGGCGCAAGAGTGGCGTGGACCAAAAGTGAAGTGGT 1542
QY 1557 CCGCGCAAGTGAACCCGCGCGCTGATCGTCACTTCAACCAACCAATGTCGCGCTGA 1616
DB 1543 CCGCGCAAGTGAACCCGCGCGCTGATCGTCACTTCAACCAACCAATGTCGCGCTGA 1602
QY 1617 TTGAACGGAACAGACCACTTGAAGCAACAGAGCCGTTGCAAGCCGATGTTCAAT 1676
DB 1603 TTGAACGGAACAGACCACTTGAAGCAACAGAGCCGTTGCAAGCCGATGTTCAAT 1662
QY 1677 TTGAACGCAACCGCGCGCTGAGAGATGACTTTGGCAAGGTCGAAGAGAGAGAGTCAAG 1736
DB 1663 TTGAACGCAACCGCGCGCTGAGAGATGACTTTGGCAAGGTCGAAGAGAGAGAGTCAAG 1722
QY 1737 AGTTCTTCCGCTGGGCGAGAGATCACTGACCGAGTGGCGCATGATTTCACTGAGAA 1796
DB 1723 ACTTTTCCGCTGGGCGAGAGATCACTGAGTGGCGCATGATTTCACTGAGAA 1782
QY 1797 AGGCTGAGGCAACAAAGACCCGCGCGATGACGCGGATTAAGCGAGGCCAAGCGG 1856
DB 1783 AGGCTGAGGCAACAAAGACCCGCGCGATGACGCGGATTAAGCGAGGCCAAGCGG 1842

QY 1857 CTTGCCCTCAGTCCGAGATCCATCGACGCTCAGACGCGGAGAGAGCTCCGTTG 1916
DB 1843 TGGCGAGTCAAGTGGGAGCCATCGACCTCAGACCGGAG---AGCTTGATCACTAG 1899
QY 1917 CCGACAGTACCAAAACAAATGTTCTGTCACGCGGAGAGCTTCAATGCTGTTCCCT 1976
DB 1900 CAGACAGTACCAAAACAAATGTTCTGTCACGCGGAGAGTCAATGCTGTTCCCT 1959
QY 1977 GCAAGCATGCGAGAGAAATGATTCGAATTTCAACTTTGCTTCAACGAGCGAGAG 2036
DB 1960 GCAAGCATGCGAGAGAAATGATTCGAATTTCAACTTTGCTTCAACGAGCGAGAG 2019
QY 2037 ACTGTTCAAGAGTCTTCCCGCGTGTGAGAAATCAACCGGT---CGTCAAGAAAGGA 2093
DB 2020 ACTGTTCAAGAGTCTTCC---CGTCAAGAAATCAACCGGT---CGTCAAGAAAGGA 2076
QY 2094 CGTATGAGAACTCTGTCGATTCATCATCTGCTGGGCGGCGCTCCGAGATTTGCT 2153
DB 2077 CGTATGAGAACTCTGTCATTCATCATCATATGAGGAAAGTGGCAGA---CGTTGCA 2133
QY 2154 CCGCTCGCATCTGCTCAACGCTGACCTGAGATGATCTGTTTCTGAGCAATTAATGACTT 2213
DB 2134 CTGCTCGCATCTGCTCAATGATGATTTGATGATGATCTTTGAAACATTAATGATTT 2193
QY 2214 AAACAGATATGAGCTGCGATGATTTCTTCCAGATTTGCTGAGAGCAACCTCTGAG 2273
DB 2194 AAATCAAGATATGAGCTGCGATGATTTCTTCCAGATTTGCTGAGAGCACTCTCTGAA 2253
QY 2274 GGCATTCGAGAGTGTGAGACTTGAACCTGAGAGCCCGCAAGCCCAACCAACGCA 2333
DB 2254 GGAATTAAGACAGTGTGAGAGCTCAAACTGGCGCCACCAACCAAGCCCGAGAGCGG 2313
QY 2334 AAGCAGAGAGAGCGCGCGGCTGCTGCTCTCTGCTCAAGATCTCGAGACCTTCAAC 2393
DB 2314 CATTAAGAGAGAGAGAGGCTTGTGCTTCTGAGTAAAGTACCTCGAGCTTCAAC 2373
QY 2394 GAGCTGCAAGAGGAGAGCGCGCTCAACGCGGAGAGAGCGAGCGCGCTCGAGAGCAAG 2453
DB 2374 GAGCTGCAAGAGGAGAGCGCGCTCAACGAGAGAGAGCGAGCGCGCTCGAGAGCAAG 2433
QY 2454 GCTTACGACAGAGCTCAAGAGGAGTGAACATCCGTAACCTGATTAACAGCGCGAG 2513
DB 2434 GCTTACGACAGAGCTCAAGAGGAGTGAACATCCGTAACCTGATTAACAGCGCGAG 2493
QY 2514 GCGAGTTTCAAGAGGCTTGAAGAAATAGTCTTTTGGGCGCAACTCGGCGAGCA 2573
DB 2494 GCGAGTTTCAAGAGGCTTGAAGAAATAGTCTTTTGGGCGCAACTCGGCGAGCA 2553
QY 2574 GTCTTCAAGAGAGAGGAGGTTCTGAACCTCTGCTGCTGATGAGAGAGCGCTAG 2633
DB 2554 GTCTTCAAGAGAGAGGAGGTTCTGAACCTCTGCTGCTGATGAGAGAGCGCTAG 2613
QY 2634 ACAGGCTCTGGAAGAAAGAGTCCGCTGAGAGAGTGCACCAAGAGCGAGACTCTCTCG 2693
DB 2614 ACAGGCTCTGGAAGAAAGAGTCCGCTGAGAGAGTGCACCTCTGAGAGCACTCTCTCG 2673
QY 2694 GCGATGCGCAAGACAGGCTGAGAGCGCTTAAAGAGAGCTCAATTTTGTGAGATGGC 2753
DB 2674 GGAACCGGAAAGAGCGGCGCGAGAGCTGCAAGAAAGATTAATTTTGTGAGATGGA 2733
QY 2754 GACTCAGAGTCAAGTCCCGATCAACCACTCTCGAGAGAGCTCGAGCAACCCCGCTGCT 2813
DB 2734 GAGCGAGCTCAAGTCACTTCAACCCAGAGCTCTGAGAGAGCTCGAGCAACCCCGCTGCT 2793
QY 2814 GTGGAGCTTACTAATGAGCTTCAAGCGGTGCGCAACCAATGCGAGCAATTAAGAGGC 2873
DB 2794 GTGGAGCTTACTAATGAGCTTCAAGCGGTGCGCAACCAATGCGAGCAATTAAGAGGC 2853
QY 2874 GCGAGAGAGTGGATTAAGCTTCAAGAAATTTGGCATTTGAGATTTCAATGCTGGCGAG 2933
DB 2854 GCGAGAGAGTGGATTAAGCTTCAAGAAATTTGGCATTTGAGATTTCAATGAGTGGCGAG 2913
QY 2934 AGAGTATCAACCAAGACCGGACCTGGGCTTGGCCACTCAATTAACCACTTAC 2993

Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGGAGCGAACCAAGTCCGCC 60
QY 61 AGACGGAGAGCTGTGCTGTGCGCGGCCCAACCGAGCGAGGAGGCGGAGAGGAGT 120
Db 61 CGACGCCCGGAGCTTTGCGCGCGGCTTCACTGAGCGAGGCGGAGGAGGAGT 120
QY 121 GCGAATCTCATCTAGAGGGTAAATCGGAGAGCGCTCCCAAGCTCCGCTCAGCGCTGA 180
Db 121 GCGAATCTCATCTAGAGGGT-----TCTGAGGAGGAGGAGTCTGA 183
QY 181 GCGAATCTCATCTAGAGGGT---GAGTGTCTGTATTAAGCTGTCACTGAGTCTTTTGC 237
Db 164 GCGAATCTCATCTAGAGGGTCTCTGTATTAAGAGTCACTGAGTGT---TTTGC 222
QY 238 GACATTTTGGAGACCAAGTGGCCATTTAGGGTATATAGCGAGAGTGAACGAGAGAT 297
Db 223 GACATTTTGGAGACCAAGTGTGTACCGTGGTATTTAAGCCGAGAGACACGAGGGT 282
QY 298 CTCGATTTTG-ACCGGCAAAATTGAAACGAGAGAGCCATGCGGGCTTCTAGAGATCG 356
Db 283 CTCGATTTTGAAGGGAGGTTTGAACGCGAGCGCATGCGGGGTTTACGAGATTG 342
QY 357 TGAATCAAGTGTCCGAGAGCCTGAGAGACACTGCGGGCATTTTGACTCGTTTGTGA 416
Db 343 TGAATTAAGTGTCCGAGAGCCTTGAACGAGCATGCGGGCATTTTGAAGCTTTGTGA 402
QY 417 GCTGGGTGGCGGAGAGAAATGGAGGCTGCCCCGATCTGAACATGGAATCTGAATCTGA 476
Db 403 ACTGGGTGGCGGAGAGAAATGGAGGTTGCGCGCAAGTCTGAACATGGAATCTGAATCTGA 462
QY 477 TTGAGCAGGAGACCCCTGACCGCTGCGCGAGAGCTGACGCGCATTTCTGTCTCAATGAC 536
Db 463 TTGAGCAGGAGACCCCTGACCGCTGCGCGAGAGCTGACGCGCATTTCTGAACGAAATGAC 522
QY 537 GCGCGGTGATTAAGGCGCGGAGGCGCTTCTTTTGTTCATGTTGAGAGAGGAGTCT 596
Db 523 GCGGTGTGATTAAGGCGCGGAGGCGCTTCTTTTGTTCATGTTGAGAGAGGAGTCT 582
QY 597 ACTTCCACCTCCATATTCGTGTGAGACACGCGGGGTCAATTCATGAGTCTGCGCGCT 656
Db 583 ACTTCCACATGACGTCTGTGTGAAACACGCGGGGTGAATTCATGAGTCTTGTGAGGTT 642
QY 657 TCTGTAGTCAATTAAGGAGCAAGCTGTGTGAGACCATCTACGCGGAGTCAAGCCGACC 716
Db 643 TCTGTAGTCAATTCGCGGAAAACTGATTCAGAGAAATTTACCGGGATGAGACCCGACTT 702
QY 717 TCGCCAACTGTTGTGCGGTGACCAAGACGCGTAATGCGCGCGGAGGAGAAACAAGTGTG 776
Db 703 TCGCCAACTGTTGTGCGGTGACCAAGAAACGAGAAATGCGCGGAGGAGAAACAAGTGTG 762
QY 777 TGAACGAGTGTCAATCCCAACTACTGCGCCCAAGACTCAAGCCCGAGCTGCAAGTGTG 836
Db 763 TGAATGAGTGTCAATCCCAACTACTGCTCCCAAAACCAAGCTGCAAGCTGCAAGTGTG 822
QY 837 CGTGTGACTAATGAGAGATATTAAGCGCTGTGTGAACCTGTGCGCGAGACGCAAAACGCG 896
Db 823 CGTGTGACTAATGAGAGATATTAAGCGCTGTGTGAATCTCAAGAGACGTAACGCT 882
QY 897 TCGTGTGCGGACACTGACCAAGTCAAGACGAGACCCGAGAGCAAGAACAGAGAAATCTGA 956
Db 883 TGTGTGTGCGGACACTGACCAAGTGTGCAAGACGAGAGAGCAAGAACAGAGAAATCTGA 942
QY 957 ACCCAATTCGTGAGGCGCTGTCAATCCGCTCAAAAACCTCGCGCGCTAATGAGAGTGTG 1016
Db 943 ATCCCAATTCGTGAGGCGCGGTGTCAATCAAAAACCTCAAGCGAGTATGAGAGTGTG 1002
QY 1017 TCGGAGTGTGAGTGTGACCGGGGACTCACTCCGAGAGAGTGTGATCAAGAGAGACGAG 1076
Db 1003 TCGGAGTGTGAGTGTGACAGGGGATTAATCTCGAGAGAGAGTGTGATCAAGAGAGACGAG 1062
QY 1077 CTTGTGTGATCTCTTCAAGCGCGCTTCAACTGTGCGGTCTCCAGATCAAGCGCTGTGTG 1136
Db 1063 CTTGTGTGATCTCTTCAATGCGCGCTCAACTGTGCGGTCTCCAAATCAAGCGCTGTGTG 1122

QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCCGACTACCTGTAGGCGCG 1196
Db 1123 ACAATGCGGCAAGATTAATGAGCTTAATAACCCGCCCTCACTCTGTGTGAGCGAC 1182
QY 1197 CTCGCGCGCGGACATTTAAACCAACCGGACTACCGGACTCTGTGAGTGAACGCTACG 1256
Db 1183 AGCCGCTGAGAGCATTTTCAGCAATCGGATTTATTAATTTTGAATCAAAACGGGTACG 1242
QY 1257 AACCTGCTTACGCGGCTCCGCTTTCTGCGCTGCGGCCAGAAAGGTTGCGGAGCGCA 1316
Db 1243 ATCCCAATATGCGGCTTCCGCTTTCTGTGAGTGGCCACGAAAGGTTGCGGAGAGGA 1302
QY 1317 ACACCATCTGGCTTTGAGGCGGCGCACACGCGGCAAGCAACATCGGAGGAGCATG 1376
Db 1303 ACACCATCTGGCTTTGAGGCTGTGCACTACCGGAGAGCAACATCGGAGGAGCATG 1362
QY 1377 CCCAGCGCGTGCCTTCTACGCGCTGCTCACTGACCAATGAGAACTTTCCTTCAATG 1436
Db 1363 CCCAGCTGTGCGCTTCTACGCGGTGTAACTGAGCAATGAGAACTTTCCTTCAAG 1422
QY 1437 ATTGCGTCCAGAAATGTGTATCTGTGTGAGAGAGGCGCAAGATGACGCGCAAGTGTG 1496
Db 1423 ACTGTGTCAAGAGATGTGTATCTGTGTGAGAGAGGAGAGATGACCGCAAGTGTG 1482
QY 1497 AGTCCGCAAGGCAATTCGCGCGGAGAGAGGAGTGTGCGTGTGACCAAAAGTCAAGTGT 1556
Db 1483 AGTCCGCAAGGCAATTCGCGGAGAGAGAGTGTGCGTGTGACCAAAATGCAAGTGTCT 1542
QY 1557 CCGCCAGATTCGACCCCAACCCCGGTGATGTGACCTTCAACCAACATGTGCGCGTGA 1616
Db 1543 CCGCCAGATTAAGACCCGATCTCCGCTGTATGTCACTTCAACCAACATGTGCGCGTGA 1602
QY 1617 TTGACGGGAAACAGACACACTTGTGAGACACAGAGCGCTTGTGAGAGCCGATGTCAAT 1676
Db 1603 TTGACGGGAACTTCAGAGACTTGTGAAACACAGAGCGCTTGTGAGAGCCGAGTGTCAAT 1662
QY 1677 TTGAATCAACCGCGCTGTGAGACATCTTTGGGCAAGGTGCAAAAGCAGAGAAATCAAG 1736
Db 1663 TTGAATCAACCGCGCTGTGAGATCATGACTTTGGAGAGGTCAACAAAGCAGAGAAATCAAG 1722
QY 1737 AGTTCTTCGCTGTGCGGAGATCAAGTACCGAGGTGCGGATGATCTACGTCAGAA 1796
Db 1723 AGTTCTTCGCTGTGCGGAGATCAAGTGTGAGGTGAGATGATCTACGTCAGAA 1782
QY 1797 AGGTGTGAGCCTCAAAAGACCGCGCCCGGATGACGCGATTAAGCGAGCCGACGCG 1856
Db 1783 AGGTGTGAGCCTCAAAAGACCGCGCCCGGATGACGCGATTAAGTGTGAGCCGACGCG 1842
QY 1857 CTTGCGCTCAGTGTGAGTCAAGTCAAGAGTCAAGAGGAGTGTGCGGTGACTTTG 1916
Db 1843 TGTGCGATCATGTTGTGAGACCAATGACGTCAAGAGCGGAG---AGCTGTGATCAATGAC 1899
QY 1917 CCGACAGTATCAAAACCAATGTCTGTCAAGCGGAGATGCTTCAAGTGTGTTCCCT 1976
Db 1900 CAGACAGTATCAAAACCAATGTCTGTCAAGTGTGAGATGATGATGCTGTTCCCT 1959
QY 1977 GCAAGACATGCGAGAGATGAATGCAATTTCAATTGCTTCAAGCAGAGAGAGAG 2036
Db 1960 GCAAGACATGCGAGAGATGAATGCAATTTCAATTTGCTTCACTCAAGAGAGAG 2019
QY 2037 ACTGTTCAAGTGTCTCCCGCGGTGTGAGATCTCAACCGGT---CGTCAAGAGAGAG 2093
Db 2020 ACTGTTCAAGTGTCTCC---CGTCAAGATCTCAACCGGTCTGTGTGTCGCAAAAGAG 2076
QY 2094 CGTATCGAAACTGTGTGCTATTCATCTGTGCGGCGGCTCCGAGATGTGTTGCT 2153
Db 2077 CGTATCGAAACTGTGTCTATTCATTCATTCATGAGAGAGAGTGTGAG---CGTGTGA 2133
QY 2154 CGGCTGTGATCTGTGCAAGTGTGACCTGTGTGATGATGATGATGATGATGATGATGAT 2213
Db 2134 CTGCGCTGATCTGTGCAATGTGTGATTTGTGATGATGATGATGATGATGATGATGAT 2193

2214 AACACGGTATGCTGCCGATGTTATCTTCAGATTGGCTCAAGACAACCTCTGAG 2273
2194 AAATCAGGATATGCTGCCGATGTTATCTTCAGATTGGCTCAAGACAACCTCTGAG 2253
2274 GGCATTGCGAGTGTGGGACTTGAACCTGAGCCCGAAGCCCAAGCCACAGCA 2333
2254 GGAATTAAGACAGTGTGGGACTCAACCTGAGCCCAACCAAGCCCGAGAGGG 2313
2334 AAGCAGACCAACGCGCCGGGTCTGTGCTTCTTGGCTCAAGTATCTTGGACCTTCAAC 2393
2314 CATTAAGACAGACAGAGGGGTCTGTGCTTCTTGGCTCAAGTATCTTGGACCTTCAAC 2373
2394 GGAATCAGCAAGAGGGGAGCCGCTCAACGCGGAGCAGAGCGGCTTGAACAGCAAG 2453
2374 GGAATCAGCAAGAGGGGAGCCGCTCAACGCGGAGCAGAGCGGCTTGAACAGCAAG 2433
2454 GCTTACGACAGAGCTCAAAAGCGGGTGAACAATCCGTAACCTGCGGTATTAACAACGCGGAC 2513
2434 GCTTACGACAGAGCTCAAAAGCGGGTGAACAATCCGTAACCTGCGGTATTAACAACGCGGAC 2493
2514 GCCGAGTTTCAAGAGCGCTCTGCAAGATAAGTCTTTTGGGGGCAACCTGGGGGAGCA 2573
2494 GCGAGTTTCAAGAGCGCTCTTAAAGATAAGTCTTTTGGGGGCAACCTGGGGGAGCA 2553
2574 GTCTTCCAGGCAAGAGCGGGTCTCGAACCTCTGCGGTCTGTTAGAGAGCGCTAG 2633
2554 GTCTTCCAGGCAAGAGGGGTCTTGAACCTCTGCGGCTGTTAGAGAGCGCTAG 2613
2634 ACGGCTCTGGAAGAAAGTCCGGTGAAGAGTGCACAAGAGCCAGACTCTCTCTCG 2693
2614 ACGGCTCTGGAAGAAAGAGGCGGTAAGAGACTCTCTGAGGAGCAGACTCTCTCTCG 2673
2694 GGCATCCGCAAGACAGGCGCAAGAGCGGCTTAAAGAGACTCTTGTGTCAGACTGAGC 2753
2674 GGAATCCGCAAGAGGCGGCAAGAGCTTCAAGAAAGATTAATTTGGTGAAGTGA 2733
2754 GACTCAGAGTACGTCCTCCGATCAACACTCTCGGAGAACTCCAGCAACCCCGCTGCT 2813
2734 GACCGAGTACGTCGTCGACCCCGAGCTCTCGGAGAGCACAAGAGCCCGCTGAGT 2793
2814 GTGGAGCTTACTACATGGCTTTCAGGCGGTGGCGCAACATGGCAGACATTAAGAGGC 2873
2794 CTGGGAATTAATACGATGCTTACAGGCGGTGGCGCAACATGGCAGACATTAAGAGGC 2853
2874 GCCGAGGAGTGGGTAATGCTCAAGAAATTTGGATTCGATTCACATGGCTGGGAGC 2933
2854 GCCGAGGAGTGGGTAATGCTTCCGAAATTTGGATTCGATTCACATGGATGGGAGC 2913
2934 AAGATCATCACACCAAGCAGCCGCACTGGGCTTGGCCACTTACATTAACACCTTAC 2993
2914 AAGATCATCACACCAAGCAGCCGCACTGGGCTTGGCCACTTACATTAACACCTTAC 2973
2994 AAGCAATCTCAAGTCTTCAACGCGGGCCAGCAAGCAACACTTACTTGGCTACAGC 3053
2974 AAGCAATCTCAAGTCTTCAACGCGGGCCAGCAAGCAACACTTACTTGGCTACAGC 3030
3054 ACCCGTGGGGGTATTTGATTTCAAGATTTCACTGCACTTTTCAACGAGTACG 3113
3031 ACCCGTGGGGGTATTTGATTTCAAGATTTCACTGCACTTTTCAACGAGTACG 3090
3114 CAGGAGCTCATCAACATTTGGGATTCGCGCCCAAGAGACTCAACTTCAACTCTTC 3173
3091 CAAGAGCTCATCAACATTTGGGATTCGCGCCCAAGAGACTCAACTTCAACTCTTC 3150
3174 AAGATCAAGTCAAGAGGTCAACGATATGCGCTCAACATATGCTTAATTACTT 3233
3151 AAGATCAAGTCAAGAGGTCAACGATATGCGCTCAACATATGCTTAATTACTT 3210
3234 ACGAGCAAGTCAAGTCTTCCGAGCTCGAGTACAGCTTCCGTAAGCTCTGAGCT 3293
3211 ACGAGCAAGTCAAGTCTTCCGAGCTCGAGTACAGCTTCCGTAAGCTCTGAGCT 3270
3294 GCGACCAAGGCTGCTCTCCGTTCCGCGGAGCTGTTATGATTTCCGCAATACGCGC 3353

3271 GCGATCAAGAGTCCCTCCGCTTCCAGCAGAGCTTCTCATGAGCAGATGA 3330
3354 TACTGACGCTCAACAATGGCAGCAGGCGGTGGAGCTTATCTTTATGCTCGAA 3413
3331 TACTGACGCTCAACAAGGAGTCAAGGAGTGGAGGCTTCTTATTTATGCTCGAG 3390
3414 TATTCCCTTCTAGATGCTGAGAAAGGCAACATTTACTTCTGACTACCTTTAG 3473
3391 TACTTCTCTCTAGATGCTGAGTCCGAAACAACTTTACTTCTGACTACCTTTAG 3450
3474 GAATGCTTTTCAACAGACTTACGCGCAGCAGCAGGCTTGAACGCGCTGATGATCT 3533
3451 GAGCTTCTTTTCAACAGACTTACGCTCAAGCAGAGTGTGAGCGCTCATGAATCT 3510
3534 CTATGACCAACAAATCTGATTTACCTGAACAGAACTCAAAATGATCCGGAATGCCAA 3593
3511 CTATGACCAACAAATCTGATTTACCTGAACAGAACTCAAAATGATCCGGAATGCCAA 3570
3594 AACAGGACTTGTGTTTACCGGTGCTCCAGCTGGCATGTCTGTCAGCCCAAAAC 3653
3571 CAGTCAAGCTTCAAGTTTCTCAGCCGAGCGAGTACATTTGGGAGCAGTCTAGGAC 3630
3654 TGGCTACTGAGCCTGTTATTCGAGCAGAGCGGCTTTTAAACAAACAGACAAAC 3713
3631 TGGCTTCTGAGCCTGTTATTCGAGCAGAGCGAGTATCAAAAGATCTGGGATTAAC 3690
3714 AACAGCAATTTTACCTGAGTGTGCTTCAAAATATTAACCTCAATGGCGCTGAATC 3773
3691 AACAGTAAATCTGAGTGTGAGTCAACCTCAATGAGCAAGCAAGCAAGCTCTG 3750
3774 ATCAACCTGAGCCTGATGCTTACCAACAAGAGCAGCAAGCAAGTCTTTCCATG 3833
3751 GTGAATCCGAGGCGCCGCAATGGCAGCAAGCAAGTGAAGAAAGTTTTCCTCAG 3810
3834 ACGGCTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAAAACCTGATTTGACAT 3893
3811 ACGGCTCATGATTTTGGAAAAAGAGCTCAGAGAAACCAATGTGAACATTTGAAG 3870
3894 GTCATGTTTCAAGCAAGAGGAAATTAAGCACTTACCTTGGCAGCAGAAATTT 3953
3871 GTCATGTTTCAAGCAAGAGGAAATTAAGCACTTACCTTGGCAGCAGAAATTT 3930
3954 GGGACCTGAGCTCAATTTCCAGAGCAGCAGCAAGCCCTGCAACCGGAGATGTGAT 4013
3931 GGTTCGTATCTCAACCTCCAGAGGCAACAGCAACACCTACCGGATGTCAAC 3990
4014 GCTATGAGCAATTTACTGCAATGTGTGCAAGATTAAGAGCTTACTGAGGCTCC 4073
3991 ACAAGAGCTTCTTCCAGGCAATGTGTGCAAGAGCAAGATGTACTTCAAGGAGCC 4050
4074 ATTTGGGCAAAATTCCTCAACAGATGACACTTTCACCCGCTCTTATGGGCGC 4133
4051 ATCTGGCAAAAGTTCACACAGAGGCAATTTTCAACCCCTCTCTCATGGGTGA 4110
4134 TTTGACTCAAGAACCGCTCTCTCATGATCTCATCAAAACAGCGCTGTCTCGAAT 4193
4111 TTGGAGCTTAAACACCTCTCTCATGATCTCATCAAGAACACCGGATCTCGAAT 4170
4194 CTTCGGGAGTTTCAAGCTCAAAAGTTTCTTCAATTCACCAATTACTCAAGAG 4253
4171 CTTCGACCACTTCAAGTGTGGCAAAAGTTTCTTCAATTCACCAAGTCTCAAGAG 4230
4254 CA-AGTATGTTGAATTTGAATGGGAGCTGCAAGAAAGAAACAGCAAGGCTGGAATCC 4312
4231 CAGGTAGGTTGAATTTGAATGGGAGCTGCAAGAAAGAAACAGCAAGGCTGGAATCC 4290
4313 CGAAGTCAATACATCAATTAATGCAAAATCTGCAAGTGAATTTTACTGAGCA 4372
4291 CGAATTAATCACTTCAACTCAACAAAGTCTGTTAATCGGAACTTACCGGTGATC 4350
4373 CAATGACTTAACTGAGCTGCGCCCATTTGCAACCGTTACTTACCGGTGAT 4432

4351 TAAATGCGTATTCAGAGCCCTGCCCCATTTGCGACAGATACCTGATCTGTGA 4410
QY 4433 ATTACGTTTAATCAATTAACCGGTTGATTCGTTTCACTTGAATTTGCTCTGCT 4492
Db 4411 ATTGCTGTATTAATCAATTAACCGGTTTATTCGTTGCACTTTGCTCTGCTAT 4470
QY 4493 TCTTATCTTATTCGTTTCAATTCGTTTATTCGTTTATTCGTTTATTCGTTT 4552
Db 4471 TTTCTTCTTATTCGTTTCAATTCGTTTATTCGTTTATTCGTTTATTCGTTT 4530
QY 4553 GATTAAGACTTATTCGTTTATTCGTTTATTCGTTTATTCGTTTATTCGTTT 4612
Db 4531 CTACAGAGG-----ACCTTATGATGAGTTGCGCATCTCTCTCTG 4574
QY 4613 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4672
Db 4575 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4634
QY 4673 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4718
Db 4635 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4680

RESULT 9
LOCUS 162303 4680 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 5658785.
ACCESSION 162303
VERSION 162303.1, GI:2480251
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4680)
AUTHORS Johnson, P.R.
TITLE Adeno-associated virus materials and methods
JOURNAL Patent: US 5658785-A 1 19-AUG-1997;
FEATURES
source 1.4680
location/Qualifiers
BASE COUNT 1198 a 1265 c 1254 g 963 t
ORIGIN

Query Match 65.3%; Score 3082.4; DB 6; Length 4680;
Best Local Similarity 80.0%; Pred. No. 0;
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGCGCACT 60
Db 1 TTGCGCACT 60
QY 61 AGACGCGAGAGCT 120
Db 61 AGACGCGAGAGCT 120
QY 121 GCGACCTCATCTAGGGGTTAATTCGGAAGCGCTCCAGCGCTGCGCGCTCAAGCGCTGA 180
Db 121 GCGACCTCATCTAGGGGTTAATTCGGAAGCGCTCCAGCGCTGCGCGCTCAAGCGCTGA 180
QY 181 CGTAAATTCATAGGAGG---GAGTGTCTCTGATTCGTTGCTGATCTGATCTCTTTGC 237
Db 164 CGTAAATTCATAGGAGGTTAGGAGGCTCTGATTCGTTGCTGATCTGATCTCTTTGC 222
QY 238 GACATTTTGGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
Db 223 GACATTTTGGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282
QY 298 CTCATTTTG-ACGCGCAAAATTTGAAGAGAGAGCATGCGCGCTTCTAGAGATG 356
Db 283 CTCATTTTGAAGAGAGGTTTGAAGAGAGAGCATGCGCGCTTCTAGAGATG 342
QY 357 TGATCAAGGTGCGAGGAGCTGAGAGAGAGCATGCGCGCTTCTAGAGATG 416
Db 357 TGATCAAGGTGCGAGGAGCTGAGAGAGAGCATGCGCGCTTCTAGAGATG 416

343 TGATTAAGTCCCGACGACCTTTGACGAGCATCTGCGCGCATTTCTGACAGCTTTGTGA 402
QY 417 GCTGGGTGCGCGGAGAGAAATGGAAGTCTGCGCGCGCATTTCTGACAGATCTGAATCTGA 416
Db 403 ACTGGGTGCGCGGAGAGAAATGGAAGTCTGCGCGCATTTCTGACAGATCTGAATCTGA 462
QY 477 TTGAGAGAGACCCCTGACCGTGGCGAGAAAGCTGAGAGCGGCTCTGATCTGAATGCG 536
Db 463 TTGAGAGAGACCCCTGACCGTGGCGAGAAAGCTGAGAGCGGCTCTGATCTGAATGCG 522
QY 537 GCGCGGTGAGTAAAGCGCGGAGCGCTCTTTCTTTGTTAGTTGAGAAAGCGAGCTCT 596
Db 523 GCGGTGAGTAAAGCGCGGAGCGCTCTTTCTTTGTTAGTTGAGAAAGCGAGAGCTCT 582
QY 597 ACTTCAACCTCCATTTCTGTTGAGAGACAGCGGAGGTTCAAAATCAGATGCTGGGCGCT 656
Db 583 ACTTCAACCTCCATTTCTGTTGAGAGACAGCGGAGGTTCAAAATCAGATGCTGGGCGCT 642
QY 657 TCTGAGTCAAGATTAGGAGCAAGCTGCTGAGACCATCTACCGCGGATCGAGCCGACCC 716
Db 643 TCTGAGTCAAGATTAGGAGCAAGCTGCTGAGACCATCTACCGCGGATCGAGCCGACCTT 702
QY 717 TGCCTCACTGTTTGGGTTGACCAAGACGCTGATGAGCGCGGAGGAGGAGCAAGCTG 776
Db 703 TGCCTCACTGTTTGGGTTGACCAAGACGCTGATGAGCGCGGAGGAGGAGCAAGCTG 762
QY 777 TGAAGAGTGTATCTATCTCCCACTACCTCTGCGGAGAGCTGAGCGGAGCTGAGTGG 836
Db 763 TGAAGAGTGTATCTATCTCCCACTACCTCTGCGGAGAGCTGAGCGGAGCTGAGTGG 822
QY 837 CGTGAATTAACATGAGAGATATTAAGGCGCTGTTTGAACCTGCGCGAGAGCAAGCGC 896
Db 823 CGTGAATTAACATGAGAGATATTAAGGCGCTGTTTGAACCTGCGCGAGAGCAAGCGC 882
QY 897 TCGTGGCGAGACCTGACCCAGCTGACGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 956
Db 883 TCGTGGCGAGACCTGACCCAGCTGACGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 942
QY 957 ACCCGAATTTCTGACGCGCTGCTGATCTGCGGTTCAAAATCTGCGCGCTGATGAGAGCTG 1016
Db 943 ACCCGAATTTCTGACGCGCTGCTGATCTGCGGTTCAAAATCTGAGCAGTATGAGAGCTG 1002
QY 1017 TCGGAGTGTGAGAGACCGGAGCATCACTCCGAGAGAGAGAGAGAGAGAGAGAGAG 1076
Db 1003 TCGGAGTGTGAGAGACCGGAGCATCACTCCGAGAGAGAGAGAGAGAGAGAGAGAG 1062
QY 1077 CCTGATCATCTCTTCAACGCGCTTCAACCTGCGGTTCCAGATCAAGGCGCTCTG 1136
Db 1063 CCTGATCATCTCTTCAACGCGCTTCAACCTGCGGTTCCAGATCAAGGCGCTCTG 1122
QY 1137 ACAATGCGGAGAGATCATGAGCTGACCAATCGCGCGGCTGATCTGTTAGGCTCCG 1196
Db 1123 ACAATGCGGAGAGATCATGAGCTGACCAATCGCGCGGCTGATCTGTTAGGCTCCG 1182
QY 1197 CTCGCGCGGAGACATTAACCAACCGCATCTACCGCATCTGAGCTGAGAGAGAGAG 1256
Db 1183 AGCGGTGAGAGACATTTCAAGCATCGATTTTAAATTTTGAATTAACGAGTACG 1242
QY 1257 AACCTGCTAAGCGCGCTCGCTCTTCTGCTGCGGCTGAGAGAGAGAGAGAGAGAG 1316
Db 1243 ATCCCAATATAGGCGCTCGCTCTTCTGAGATGAGGCAAGAGAGAGAGAGAGAGAG 1302
QY 1317 ACAACATCTGAGCTGTTGGGCGGAGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAG 1376
Db 1303 ACAACATCTGAGCTGTTGGGCGGAGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAG 1362
QY 1377 CCCAGCGGTGCGCTCTACGCGCTGCTCAACCTGAGACATGAGAGAGAGAGAGAGAG 1436
Db 1363 CCCAGCGGTGCGCTCTACGCGGTGCTCAACCTGAGACATGAGAGAGAGAGAGAGAG 1422
QY 1437 ATTGCTGACCAAGATGATGATCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
Db 1423 ACTGCTGACCAAGATGATGATCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482

QY	1457	AGTCGCCCAAGGCCATTCTTCGGCGGCGACCAAGTGGCGGTGGACCAAAAGTGCAGTCT	1558
Db	1483	AGTCGGCCCAAGGCCATTCTTCGGAGGAACCAAGTGGCGGTGGACCAAAAGTGCAGTCT	1542
QY	1557	CGGCCCAATGCAGCCCCCAACCCCGGTGATCGTCACTCCACAACCAATGTGGCCGCTGCA	1616
Db	1543	CGGCCCAATGATACCCGACTCCCGGTGATCGTCACTCCACAACCAATGTGGCCCGTGA	1602
QY	1617	TTGACGGGAACGACACCACTTTCGAGCACACAGACGCCGTTGCAGGACCGGATGTTCAAT	1676
Db	1603	TTGACGGGAACCTCAACGACCTTTCGAAACACAGCACGCCGTTGCAGAGCCGGATGTTCAAT	1662
QY	1677	TTGAATCTACCCCGCGTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAG	1736
Db	1663	TTGAATCTACCCCGCGTGGAGCATGACTTTGGGAAGGTGACCAAGCAGGAAGTCAAG	1722
QY	1737	AGTCTTTCGGCTGGGCGCAGAGATCAAGTACCGAGGTGGCGCATGATGTTCTACGTCAAGA	1796
Db	1723	ACTTTTTCGGGTGGGCGCAAGATCAAGTGTGGTGAAGTGAATTTCTACGTCAAAA	1782
QY	1797	AGGCTGAGCCCAACAAAGATCCCGCCCGGATGATCGCGATTAAGCGACCCCAAGCGGG	1856
Db	1783	AGGCTGAGCCCAACAAAGATCCCGCCCGGATGATCGCGATTAAGTGAAGCCCAACCGGG	1842
QY	1857	CGTGGCCCTCAAGTGGGGATCCATGAGAGTCAAGCGGGAAGGAGCTCGGTGGACTTTGG	1916
Db	1843	TGCGGAGTCAAGTGGCGAGCCATGAGAGTCAAGCGGGAAGGAGCTCGGTGGACTTTGG	1899
QY	1917	CCGACAGGATCCAAACAAATGTTCTCGTCAAGCGCGGATGTTGATGCTGTGTTCCCT	1976
Db	1900	CAGACAGGATCCAAACAAATGTTCTCGTCAAGTGGGATGATGTAGTGTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATCAAGATTTCAACTTTCGTTCAACGACCGGACAGAG	2036
Db	1960	GCAGACATGCGAGAGATGAATCAAGATTTCAAAATTCGTTCACTCAACGACAGAGAG	2019
QY	2037	ACTGTTCAAGTGCCTCCCGGCGGTGAGATCTCAACCGGT---CGTCAAGAAAGAGA	2093
Db	2020	ACTGTTTGAAGTGCCTTTC---CGTGTCAAGATCTCAACCGGTTTCTGTCTGCAAAAAG	2076
QY	2094	CGTATCGGAACCTCTGTGCATTCATCACTGCTGGGGCGGGCTCCGAGATTTGCTTCT	2153
Db	2077	CGTATCAGGAACCTGTGTCTACATTCATCACTGATGGGAAGGTGCAAG---CGTGTCA	2133
QY	2154	CGGCTCGGATTTGGTCAACGTGGAACCTGGATGACTGTGTTCTGAGCAATAATGACTT	2213
Db	2134	CTGCTCGGATTTGGTCAATGTGGAATTTGGAGACTGATCTTTGAAACAAATAATGACTT	2193
QY	2214	AAACAGGATGGCGCGGATGGTATCTTCAGATTGGCTCAAGACAACCTCTGAG	2273
Db	2194	AAATCAGGATGGCTGGCGGATGGTATCTTCAGATTGGCTTGAAGACAATCTCTCTGA	2253
QY	2274	GGCATTCGGAATGTGGGACTTGAACCTTGAAGCCCGGAACCCCAAGCCAACCAAGCAA	2333
Db	2254	GGAATTAAGCAATGTGTGGAAGTCAAACTTGGCCCAACCAACCAACCCGCAAGCGG	2313
QY	2334	AAGCAGAGACAGCGCCGGGGTGTGGTGTCTTCTGGCTACAAATTAATCTGGACCTTTCAC	2393
Db	2314	CATTAAGACAGACAGAGGGGGTGTGTGTCTTCTGGGTACAAATTAATCTTGGACCTTTCAC	2373
QY	2394	GGACTTCGACAAAGGGGGAGCCGCTCAACGCGGCGAGACGACGAGCGCCCTGAGACGACAAG	2453
Db	2374	GGACTTCGACAAAGGGGGAGCCGCTCAACGAGGCGAGACGCGCGGCCCTGAGACGACAAG	2433
QY	2454	GGCTTACGACCAAGCACTCAAGCGGGGTGAACAATCCGTACCTGCGGTATTAACAAGCGCGAC	2513
Db	2434	GGCTTACGACCGGCACTGACAGCGGGAACAACCCGTACTCAATTAACAACAAGCGCGAC	2493
QY	2514	GCGGAGTTTCAGGAGCGTCTGCAAGAATAGTCTTTTGGAGGCAACTCGGGGAGACA	2573
Db	2494	GCGGAGTTTCAGGAGCGCTTAAAGAAATAGTCTTTTGGAGGCAACTCGGAGAGACA	2553

OY	2574	TTCTTCCAGGCGAAGAGCGGGTTCTGAAACTCTCGGTCTGGTTTGAAGAAAGGCGCTAAG	2653
Db	2554	GTCTTCCAGGGAAAAAGAGGGTTCTTGAACCTCTGGGCTGGTGGAACTCTTTAAG	2613
OY	2634	ACGGCTCTCTGGAAAAGAAACGTCGCGTGAAGACAGTCGCGACAAGAGCCAGACTCTCTCTCG	2693
Db	2614	ACGGCTCCGGGAAAAAAAAGAGGCGGTGAAGACACTCTCTGTGGAGCCAGACTCTCTCTCG	2673
OY	2694	GGCATTCGGCAAGACAGGCGACAGGCGCGCTTAAAAAGAGCTCAATTTTGGTCAAGCTGGC	2753
Db	2674	GGAACTCGGAAAAAGGGGGCGCCAGAGACCTCGCAAGAAAAAGATTGAATTTTGGTCAAGACTGGA	2733
OY	2754	GACTCAGAGTCAGTCCCCGATTCACAACACTCTCGAGAAACCTCCAGCAACCCCGCTGCT	2813
Db	2734	GAACCAAGCTCAAGTACTGAACCCCAACCTCTGCGAAGAGCAACAGAGCCGCCCTCTGCT	2793
OY	2814	GTGGAGCCTTACTACATATGGCTTTACAGCGGTGGCGACCAATATGCGACAAATTAAGAAAGC	2873
Db	2794	CTGGGAACTAATATGAATGGCTTACAGGCGATGGCGCACCAATATGGCGACAAATTAACGAGGGC	2853
OY	2874	GGCGACGGAGTGGGTAAATNGCTCGAGAAATTTGGCATTTGCGATTTCAATGAGCTGGGGAGC	2933
Db	2854	GGCGACGGAGTGGGTAAATTTCTCCGAAATTTGGCATTTGCGATTTCAATGAGTGGGGAGC	2913
OY	2934	AGAATCATCAACACAGACACCCGCACTCTGGGCTTTGGCCACTTAACAATTAACAATCTTAC	2993
Db	2914	AGAATCATCAACACAGACACCCGCACTCTGGGCTCTGGCCACTTAACAACCACTCTTAC	2973
OY	2994	AAGCAATATTCACATATGCTTCAACGGGGGGCCAGACAAGAACCAATCACTTCCGCTAAGC	3053
Db	2974	AAACAATATTCACACCAATCA - - GGAGCCTCGAAGCAATCACTACTTTGGCTTACAGC	3030
OY	3054	ACCCCTGGGGGTATTTTGAATTTTCAGACAGATTCACATGCACTTTTACACACGTAAGTG	3113
Db	3031	ACCCCTGGGGGTATTTTGAATTTTCAACAGATTCACATGCACTTTTACACACGTAAGTG	3090
OY	3114	CAGGACTCATCAACAACAATTTGGGGATTTCCGGGCCAAGAGACTCAACTTCAAACTTTTC	3173
Db	3091	CAAGACTCATCAACAACAATTTGGGGATTTCCGACCCAAAGAGACTCAACTTCAAACTTTT	3150
OY	3174	AACATCCAAATGCAAGAGGTCAACAACAATATGATGGGCTCAACATTCGCTAATTAACCTT	3233
Db	3151	AACATTCAAATGCAAGAGGTCAACAACAATATGATGGGCTCAACATTCGCTAATTAACCTT	3210
OY	3234	ACGAGCAACGATTCAGTCTTCTCGAGCTCGAGTACCAAGCTTCCGTACGTTCTCGGCTCT	3293
Db	3211	ACGAGCAACGATTCAGTCTTCTCGAGCTCGAGTACCAAGCTTCCGTTACGTTCTCGGCTCTG	3270
OY	3294	GCGACCAAGGCTGCTCCCTCCGCTTCCCGCGGAGCGTTTCAATGATTTCCGCAATACAGGC	3353
Db	3271	GCGACCAAGGATGCTCCCGCGCTTCCCGCGGAGCGTTTCAATGATTTCCGCAATATAGCA	3330
OY	3354	TACCTGAACGCTCAACAATATGGACGCAAGCCGTGGGAAGTTTATCTTTTACTGCTGAAA	3413
Db	3331	TACCTGAACCTGAAACAACGGGAGTCAAGGAGTGAAGCGCTTCTTATTTTACTGCTGAG	3390
OY	3414	TATTTCCCTTCTCAGATGCTGAGAAACGGGCAACAATTTAACCTTCACTCACTCACTTTGAG	3473
Db	3391	TACTTTCTTCTCAGATGCTGCGATGACGGAAACAATTTAACCTTCACTCACTCACTTTGAG	3450
OY	3474	GAAATGCTTTTCCACAGACAGTATGAGCCACAGCCAGAGCTTGAACCGGCTGATGAATCT	3533
Db	3451	GAAATGCTTTTCCACAGACAGTATGCTTCAAGCCAGAGTCTGAGACCGTCTCATGAATCT	3510
OY	3534	CTCATTCAGCAATATCTGTATTTACTTGAACGAATCTAATAATCACTCCGGAGTCCCA	3593
Db	3511	CTCATTCAGCAATATCTGTATTTACTTGAACGAATCTTCAAGTGAACCAACACAGC	3570
OY	3594	AAACAAGACTTGCATTTAAGCCGATGGGTCTCACTGGCAATGTCTGTTAGCCCAAAAAC	3653
Db	3571	CAATCAAGGCTTCAAGTTTCTCAAGCCGAGACAGTACATTTCCGGAGCAAGTCTAGAAC	3630
OY	3654	TGGTACTCGAACCTGTATTTGCGACACAGCGCTTTCTAACAACAAAAACAGACAACAC	3713

D	b	3631	TGCTTCTTGGACCTGTATACCGCAGACGAGATTAAGACATCTGCGGATTAACAC	3630
Q	y	3714	AACAGCAATTTTAACTGTGACTGTGTCTTCAAAATATAACTCAATGGCGTGAATCAATC	3773
D	b	3691	AACAGTGAATCTCGTGGACTGTAGACTACCAAGTATCCACTCAATGGCAGAGACTCTCTG	3750
Q	y	3774	ATCAACCTTGGCAGCTGCTATGGCCCTCAACAAGACGACGAAGCAAGTTCTTTCCATG	3833
D	b	3751	GTAATCCGGGGCCCGCATGTGCACCAAGGACGATGAAGAAATTTTCTTCAG	3810
Q	y	3834	AGCGGTGCATGATTTTGGAAAAAGAGACCGCCGAGCTTCAAACTGCAATTGSGACAT	3893
D	b	3811	AGCGGGGTTCATCTTTTGGAGAGCAGGCTCAGAGAAAACAAATGTGAACATTGAAAG	3870
Q	y	3894	GTCAATTAACAGACGAAGAGAAATTAAGCCATAACCTGTGGCCACCGAAAGATT	3953
D	b	3871	GTCAATTAACAGACGAAGAGAAATGGAACAACCAATCCGTTGGCTACGAGAGATAT	3930
Q	y	3954	GGGACCGTGGCAGTCAATTTCCAGACGACGACACACCTCGGACCTGGAGATGTGAT	4013
D	b	3931	GGTTCGTATCTTACCAACCTCCAGAGAGCAACAGCAAGAGCTTACCGCAGATGTCAAC	3990
Q	y	4014	GCTATGGAGCATTAACCTGGCATGGTGTGGCAAGATAGAGCGTGTACTGTGAGGATGCC	4073
D	b	3991	ACACAGGCGTCTTCCAGGCAATGTCTGTGGCAGACAGAGATGTGTACTTCAAGGAGCC	4050
Q	y	4074	ATTGGGCGAAATTCCTCCACACAGATGACATTTCAACCGCTCTCTCTTATGGGCGGC	4133
D	b	4051	ATCTGGGCAAGATTCACACACGAGCGGACATTTTCAACCCCTCTCCCTCATGGGTGA	4110
Q	y	4134	TTTGGACTCAAGAACCGCGCTCTCTCAATCTCTCATCAAAACAGCGCTGTCTCTGCAAT	4193
D	b	4111	TTTGGACTTAAACACCTCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACTGTGCAAT	4170
Q	y	4194	CCTCCGGCGAGTTTTCAGTACATAAAAGTTTGCTTCAATTCATCAACCCCAATATCTCACAGGA	4253
D	b	4171	CCTTCGACCACTTCAGTGTGGCAAAAGTTTGCTTCTTCATCAACAGATATCTCAACGGGA	4230
Q	y	4254	CA-AGTAGTGTGAATTTGAATGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGGAATCC	4312
D	b	4231	CACGTCAGCGTGGAGATCGATGGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGGAATCC	4290
Q	y	4313	CGAAGTGAATGACATCAATTTATGCAAAATCTGCCAAGTTATTTTACTGTGGACAA	4372
D	b	4291	CGAAATTCAGTACCTTCCAACTACACAGATCTGTATGTGTGAGACTTACCGTGAATAC	4350
Q	y	4373	CAATGACCTTATATCTGAGCGCTGCGCATTTGGACACCGCTTACCTTACCGTCCCTGTA	4432
D	b	4351	TAAATGGCGTATTCAGAGCCTCGCGCCATTGGACACAGATTAACCTGACTGTGAATCTGTA	4410
Q	y	4433	ATTACGTGTATCAATTAACCGGTGTGAATGTTCAGTTGAACTTTGCTCTCTGTCCT	4492
D	b	4411	ATTGCTGTATATCAATTAACCGTTTAATGTGGCAGTTGAACCTTTGGTCTCTGCGTAT	4470
Q	y	4493	TCTTATCTTATCGGTTACCAAGTTTATAGCTTTACATTAACGTGCTGTGCGCTGCC	4552
D	b	4471	TTCTTTCTTATCTAGTTTTCATGTGCTACGTGATTAATTAAGCATGGCGGCTTAACTATTA	4530
Q	y	4553	GATTAAGACTTAACTGTCATCGGGTTAACCCATAGATGAGATGGCCACTCCCTCTGAC	4612
D	b	4531	CTACAAAGG-----ACCCTAATGATGAGATTGGCCACTCCCTCTCTGAC	4574
Q	y	4613	GCCTGTGCTGCTCGGTGGGGCTGTGGACCAAAAGTCCGAGACGCGAGACTCTGTGC	4672
D	b	4575	GCCTGTGCTGCTCACTGAGGCGGGGACCAAAAGTGTGCGCGAGCGCGGGCTTTGGCC	4634
Q	y	4673	TGCGGGCCCCACGAGCGAGCGTGGCCACAGAGGAGGTGGGCA	4718
D	b	4635	GGGGGGCTTCAGTAGCGAGCGCGCCACAGAGGAGGTGGCCAA	4680

AX135805	LOCUS	AX135805	Sequence 5 from Patent WO0132711.	4675 bp	DNA	Linear	PAT 29-MAY-2001
AX135805	DEFINITION	AX135805	Sequence 5 from Patent WO0132711. <td>4675 bp</td> <td>DNA</td> <td>Linear</td> <td>PAT 29-MAY-2001</td>	4675 bp	DNA	Linear	PAT 29-MAY-2001
AX135805	ACCESSION	AX135805					
AX135805	VERSION	AX135805.1	GI:14272040				
AX135805	KEYWORDS						
AX135805	SOURCE						
AX135805	ORGANISM						
AX135805	REFERENCE						
AX135805	AUTHORS						
AX135805	TITLE						
AX135805	JOURNAL						
AX135805	FEATURES						
AX135805	source						
AX135805	CDS						
AX135805	ORIGIN						
AX135805	BASE COUNT						
AX135805	Query Match						
AX135805	Best Local Similarity						
AX135805	Matches 3779;						
AX135805	Conservative						
AX135805	Mismatches						
AX135805	Indels						
AX135805	Gaps						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels						
AX135805	Gaps						
AX135805	Length						
AX135805	Score						
AX135805	Pred. No. 0;						
AX135805	Indels		</				

477 TTGAGAGGCAACCCCTGACCGGTGCGGAGGCTGACGCGGACTTCTGTCGAATGAC 536
463 TTGAGAGGCAACCCCTGACCGGTGCGGAGGCTGACGCGGACTTCTGTCGAATGAC 522
537 GCCCGGTGAGTAAAGGCGCGGAGGCGCTCTTCTTGTGTCAGTTCAGAGAGGCGAGTCT 596
523 GCCGTGAGTAAAGGCGCGGAGGCGCTCTTCTTGTGTCAGTTCAGAGAGGCGAGTCT 582
597 ACTTCCACTTCATATTTGAGTGAAGACACGCGGCGTCAATTCATGTCGTCGCGCGCT 656
583 ACTTCCACTTCAGTGTCTGAGAAACACCGGCGTCAATTCATGTCGTCGCGCGCT 642
657 TCTGAGTCAAGTAAAGGCAAGCTGAGACACATCTACCTGCGGAGTCGAGCGGACCC 716
643 TCTGAGTCAAGTAAAGGCAAGCTGAGACACATCTACCTGCGGAGTCGAGCGGACCC 702
717 TGCCCAACTGCTTGGCGGTGACCAAGACGCGTAAATGCGCGGAGGAGGAAAGAGTGG 776
703 TGCCCAACTGCTTGGCGGTGACCAAGACGCGTAAATGCGCGGAGGAGGAAAGAGTGG 762
777 TGAACAGTGTCTAATCCCACTAATCTCTGCGGAGTCAAGCTGAGTGAAGTGG 836
763 TGAATGAGTGTCTAATCCCACTAATCTCTGCGGAGTCAAGCTGAGTGAAGTGG 822
837 CGTGAAGTAAAGGAGATATTAAGCGCTGTTGAACCTGCGGAGGCGCAAGCGC 896
823 CGTGAAGTAAAGGAGATATTAAGCGCTGTTGAACCTGCGGAGGCGCAAGCGC 882
897 TCGTGGCGCAGACCTGACCCAGCGTCAAGCGGAGGAGGAGGAGATCTGA 956
883 TCGTGGCGCAGACCTGACCCAGCGTCAAGCGGAGGAGGAGGAGATCTGA 942
957 ACCCAATTCGAGCGCGCTGTCATCCGCTCAAAAACCTCGCGCGTCAATGAGAGTGG 1016
943 ATCCCAATTCGAGCGCGCTGTCATCCGCTCAAAAACCTCGCGCGTCAATGAGAGTGG 1002
1017 TCGGAGTGTGAGTGAACCGGAGGAGTCACTCCGAGAGGAGTGAATCCAGAGAGGAGG 1076
1003 TCGGAGTGTGAGTGAACCGGAGGAGTCACTCCGAGAGGAGTGAATCCAGAGAGGAGG 1062
1077 CTTGCTGATCTCTTCAACGCGCGCTTCAACCTGCGGCTCCAGATCAAGCGCTCTGG 1136
1063 CTTGCTGATCTCTTCAACGCGCGCTTCAACCTGCGGCTCCAGATCAAGCGCTCTGG 1122
1137 ACAATGCGGAGATATGAGCGCTGACCAATCCGCGCGGAGTCACTGCTGAGCGCGG 1196
1123 ACAATGCGGAGATATGAGCGCTGACCAATCCGCGCGGAGTCACTGCTGAGCGCGG 1182
1197 CTTGCGCGCGGAGATATTAAGCAACGCGATCTACCGGATCTTGAAGTGAACGAGTACG 1256
1183 AGCCCGGTGAGGAGATTTCCAGCAATCGGATTTTAAATTTTGAAGTAAACGAGTACG 1242
1257 AACCTGCTTACGCGCGCTCGGTCTTCTCGGCTGAGGCGCAAAAAGTTCGGAGAGCGCA 1316
1243 ATCCCAATATAGCGGCTTCCGCTCTTCTGAGGATGAGGCGCAAAAAGTTCGGAGAGG 1302
1317 ACAATGCTGCTGCTTTGGCGCGGAGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1376
1303 AGACCAATCTGCTGCTTTGGCGCGGAGCAACGAGGAGGAGGAGGAGGAGGAGGAGG 1362
1377 CCGCAGCGGTGCGCTTCTAAGGCTGCGTCAATGAGTGAACCTTCTTCAATG 1436
1363 CCGCAGCGGTGCGCTTCTAAGGCTGCGTCAATGAGTGAACCTTCTTCAATG 1422
1437 ATTGCGTGAAGAGATGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1496
1423 ACTGTCGAGCAAGATGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1482
1497 AGTCCGCAAGGAGGATCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1556
1483 AGTCCGCAAGGAGGATCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1542

1557 CCGCCAGATGACCCCAACCCCGGTGATGTCACCTTCAACCAACCAATGAGCGCGTGA 1616
1543 CGGCCAGATGACCCCAACCCCGGTGATGTCACCTTCAACCAACCAATGAGCGCGTGA 1602
1617 TTGACGGGAAACAGACCACTTTCAGACCAAGAGCGGTTTGACGAGACCGGATTTCAAT 1676
1603 TTGACGGGAACTCAAGACCTTTCAGACCAAGAGCGGTTTGACGAGACCGGATTTCAAT 1662
1677 TTGACCTACCCCGGTGAGGAGATGACCTTTGGCAAGGAGGAGGAGGAGGAGGAGG 1736
1663 TTGACCTACCCCGGTGAGGAGATGACCTTTGGCAAGGAGGAGGAGGAGGAGGAGG 1722
1737 AGTTCTTCCGCTGAGGAGATGACGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1796
1723 ACTTTTTCGCTGAGGAGGAGATGACGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAG 1782
1797 AGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1856
1783 AGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1842
1857 CTTGCCCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1916
1843 TCGCGAGTCAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1899
1917 CCGACAGGATCAACCAATGTTCTGTCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1976
1900 CAGACAGGATCAACCAATGTTCTGTCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1959
1977 GCAAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2036
1960 GCAAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2019
2037 ACTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2093
2020 ACTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2076
2094 CGTATCGAAGATCTGTCGATTCATCATCTGTCGAGGAGGAGGAGGAGGAGGAGGAGG 2153
2077 CGTATCGAAGATCTGTCGATTCATCATCTGTCGAGGAGGAGGAGGAGGAGGAGGAGG 2133
2154 CGGCTGCGATCTGTCGATTCATCATCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2213
2134 CTTGCTGCGATCTGTCGATTCATCATCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2193
2214 AAACAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2273
2194 AAATCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2253
2274 GGCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2333
2254 GGAATTAAGACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2313
2334 AAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2393
2314 CATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2373
2394 GGAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2452
2374 GGAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2433
2453 GGCCTACGACCAAGCAGCTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2512
2434 AGCCTACGACCAAGCAGCTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2493
2513 CGCGAGGATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2572
2494 CGCGAGGATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2553
2573 AGTCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2632
2554 AGTCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2613
2633 GACGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2692

Db 2614 GAGGCGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTC 2673
Qy 2693 GGGCATCGGGAAGACAGGCGGAGCGCCGCTAAAAAGAGACTCAATTTTGGTCAAGCTCG 2752
Db 2674 GGGAGCCGGAAGAGCGGGCCAGAGCTGCAAGAAAAAGATTGATTTTGGTCAAGCTCG 2733
Qy 2753 CGACTCAGAGTCAAGTCCCGATCCACCACTCTCGAGAACTCTCGAGCAACCCCGCTCG 2812
Db 2734 AGACGCAAGTCAAGTACTTACCCCGAGCTCTCTCGAGAGCCACAGAGAGCCCTCTCG 2793
Qy 2813 TGTGGAGCTACTACAAATGGCTTCAGGCGGTGGGCGCAACATGGCAAGACAAATACGAAG 2872
Db 2794 TGTGGAGCTAAATACGATGGCTACAGGCAATGGGCGCAACATGGCAAGACAAATACGAAG 2853
Qy 2873 CGCGGAGCGAGTGGGTAAATGCTTCAGAGAAATTTGGCATTTGCCATGGCTGGCGCA 2932
Db 2854 CGCGGAGCGAGTGGGTAAATGCTTCAGAGAAATTTGGCATTTGCCATGGGTGGCGCA 2913
Qy 2933 CAGAGTCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2992
Db 2914 CAGAGTCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2973
Qy 2993 CAGAGCAATCTCAGTGGCTTCAAAGGCGGCGCAGCAAGCAACCACTACTTGGCTACAG 3052
Db 2974 CAAACAAATTTCCAGCCCAATCA---GGAGCCTGGAAGCAATCACTACTTGGCTACAG 3030
Qy 3053 CAGCCCTGGGGGTATTTTGAATTTTCAAGATTTCACTGCACTTTTCACTGCACTG 3112
Db 3031 CAGCCCTGGGGGTATTTTGAATTTTCAAGATTTCACTGCACTTTTCACTGCACTG 3090
Qy 3113 GAGCGAGCTCATCAACAAATTTGGGGAATTCGGGCGCAGAGCACTCACTCACTCACT 3172
Db 3091 GCAAGAGCTCATCAACAACTGGGGAATTTCCAGCCAGAGACTCACTTCAAGCTCTT 3150
Qy 3173 CAAATCTCAAGTCAAGAGAGTCAAGAGATTTGGGCTCAACCACTCACTCACTCACT 3232
Db 3151 TAAATTTCAAGTCAAGAGAGTCAAGAGATTTGGGCTCAACCACTCACTCACTCACT 3210
Qy 3233 TACAGCAGGTTCAAGTCTTCTCGAGACTGGAGTACAGCTTCCGTAGCTCTCGGCTC 3292
Db 3211 TACAGCAGGTTCAAGTCTTCTCGAGACTGGAGTACAGCTTCCGTAGCTCTCGGCTC 3270
Qy 3293 TGGGCAACGAGGCGCTCCCTCGGTTCCGGGCGAGCTGTATGATTCGGCAATACGG 3352
Db 3271 GGGGCAATCAAGATGCTCTCCCGCTTCCAGAGAGCTCTTCAATGTTGCAAGATGG 3330
Qy 3353 CTACCTGACGCTCAACAAATGGAGCCAGGCGGTGGGAGCTTCACTCTTTTACTGCTGGA 3412
Db 3331 ATACCTCAACCTGAACAAAGGAGTCAAGGATGAGACGCTCTTCAATTTTACTGCTGGA 3390
Qy 3413 ATATTTTCTCTTCAAGTGTGAGAAAGGGGCAACACTTTTCACTTCACTGCACTTTGA 3472
Db 3391 GATCTTCTCTTCAAGTGTGAGAAAGGGGCAACACTTTTCACTTCACTGCACTTTGA 3450
Qy 3473 GGAAGTGCCTTTCACAGCAGTACGCGACAGCGAGGCTGGAACGGGTGATGATCC 3532
Db 3451 GGAAGTGCCTTTCACAGCAGTACGCGACAGCGAGGCTGGAACGGGTGATGATCC 3510
Qy 3533 TCTCATGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAAGTCCGGAAGTCCCA 3592
Db 3511 TCTCATGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAAGTCCGGAAGTCCCA 3570
Qy 3593 AAAACAAGACTTGTCTGTATAGCGGTGGGTTCAGCTGGGATGTCTGTTCAGCCCAAAA 3652
Db 3571 GCAAGTCAAGGCTTCAGTTCCTCAGGCGGAGGAGGATTCATTCGGGACCAAGTCTAGGAA 3630
Qy 3653 CTGGCTACCTGAGCCCTGTATATGGGAGAGCGGCTTTCTAAAAACAAGACAGCAACAA 3712
Db 3631 CTGGCTACCTGAGCCCTGTATATGGGAGAGCGGCTTTCTAAAAACAAGTCTAGGAA 3690
Qy 3713 CAAACAGCAATTTTAACTGAGCTGTGCTTCAAAATTAACCTCAATGGGCGTGAATTCAT 3772

Db 3691 CAAAGTGAATACTCGTGAAGTGAAGTCAACAGTACCACTCAATGAGAGAGACTCTCT 3750
Qy 3773 CATCAACCTCTGGAGCTGTATGGCTTCAACAAAGAGAGAGAAAGTCTTTTCCAT 3832
Db 3751 GGTGAATC-----CGGCAATGGCAAGCAAGAGAGATGAAGAAAGTTTTCCTCA 3804
Qy 3833 GAGCGGTGCATGATTTTGGAAAAAGAGCGCGGAGCTTCAACACTGTCATTTGACAA 3892
Db 3805 GAGCGGTGTCTCATCTTTTGGAGAGAGAGCTTCAAGAAAAACAATGTGAACCTTGA 3864
Qy 3893 TGTATGATTAACAGAGAGAGAGAAATTAAGCACTAACCTGTGGCCACCGAAAGATT 3952
Db 3865 GGTCAATGATTAACAGAGAGAGAGAAATTAAGCACTAACCTGTGGCTACAGAGAGTA 3924
Qy 3953 TGGAGACCTGGAGAGTAAATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4012
Db 3925 TGGTCTGTATTAACCAACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3984
Qy 4013 TGTATGAGAGCATTAACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4072
Db 3985 CACAG 4044
Qy 4073 CATTTGGGCGCAAAATTTCTCAACAGATGAGACATTTCAACCGCTCTCTTATAGGAG 4132
Db 4045 CATCTGGGCGCAAAATTTCTCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4104
Qy 4133 CTTTGAGATCAAGAAACCGGCTCTCAAGATCTCTCAATCAAAACAGAGAGAGAGAG 4192
Db 4105 ATTGAGATCAAAACAG 4164
Qy 4193 TCTTCCGGGAGAGTTTTCAGATCAAGAGTTTGGCTTCAATCATCAACCAATACCTCA 4252
Db 4165 TCTTTCAG 4224
Qy 4253 ACA-AGTGAAGTGAAGAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4311
Db 4225 ACAAG 4284
Qy 4312 CCGAAGTGAAGTGAAG 4371
Db 4285 CCGAAGTGAAGTGAAG 4344
Qy 4372 ACAATGAGTGAAG 4431
Db 4345 CTATAG 4404
Qy 4432 AATTAGAGTGAAG 4491
Db 4405 AATTAGAGTGAAG 4464
Qy 4492 TTTTATCTTATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4551
Db 4465 TTTTATCTTATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4524
Qy 4552 CGATAAAGAGTGAAG 4611
Db 4525 ACTACAGAG-----ACCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4568
Qy 4612 CGGAGTGTCTGCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4671
Db 4569 CGGAGTGTCTGCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4628
Qy 4672 CTGGCGGCGCCACAG 4718
Db 4629 CGGCGGCGCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4675

RESULT 11
AX286292
LOCUS AX286292 4675 bp DNA linear PAT 21-NOV-2001
DEFINITION Sequence 1 from Patent WO0180840.
ACCESSION AX286292
VERSION AX286292.1 GI:17048540

Db 1843 TGGCGAGTCAGTTGCGGACCATCGAGCTCAGACGCGGA---AGCTTCATCAACTACG 1899
Qy 1917 CGGACAGGTACAAAAGAAATGTTCTGTGACCGGGGACGTTCAGATGCTGTTCCCT 1976
Db 1900 CAGACAGGTACAAAAGAAATGTTCTGTGACCGGGGACGTAATCTGATCTGTTCCCT 1959
Qy 1977 GCAAGCATCGAGAGATGATCAGAAATTCACATTTGCTTCACGACGAGGACGAG 2036
Db 1960 GCAGACAAATCGAGAGATGATCAGAAATTCACATTTGCTTCACGACGAGAG 2019
Qy 2037 ACTGTTCAAGTGTCTCCCGGGGTGTCAGAAATCTCAACCGGT---CTGCAAAAAGGA 2093
Db 2020 ACTGTTCAAGTGTCTCC---CGTGTCAAAATCTCAACCGGTCTGTGTCAAAAAGG 2076
Qy 2094 CGTATCGGAATCTGTGCAATTCATCATGCTGGGGGGGGCTCCGAGATGCTTGTCT 2153
Db 2077 CGTATCGGAATCTGTGCAATTCATCATGGAAGAGGAGGCGA---CGCTTGCA 2133
Qy 2154 CGGCTGCGATCTGTTCAGATGACCTGAGTGAATGCTGTTTCTGAGCAATTAATGACTT 2213
Db 2134 CTGCTGCGATCTGTTCAGATGATGATTTGATGATGATCTGATCTTTGAAATTAATGATTT 2193
Qy 2214 AAACAGATATGCTGCTCGATGTTATCTTCAGATTTGCTCGAGCAACTCTCTGAG 2273
Db 2194 AAATCAGTATGCTGCTCGATGTTATCTTCAGATTTGCTCGAGGACACTCTCTGAA 2253
Qy 2274 GGCATTCGCAATGCTGAGCTTGAACCTGAGCCCGGAAAGCCCAAGCCCAACGACAA 2333
Db 2254 GGAATAGACAGTGTGAGAGCTCAACCTGAGCCCAACGACCAAGCCCGGAGACCGG 2313
Qy 2334 AAGCAGGACAGCGGCGGGGTCTGTGCTTCTGTGCTACAGTACCTCGAGCCCTTCAAC 2393
Db 2314 CATAGAGACAGCAGAGGGGTCTGTGCTTCTGTGCTACAGTACCTCGAGCCCTTCAAC 2373
Qy 2394 GGAATCAGCAAGGGGAGGCGGCTCAACGCGCGGAGCGAGCGGCGCTCGAGACG-ACAA 2452
Db 2374 GGAATCAGCAAGGGGAGGCGGCTCAACGAGGAGACGCGGCGCTCGAGACGTAACA 2433
Qy 2453 GGCCTGACAGCAGCTCAAGCGGCTCAACCTGCTACCTGCGGTATTAACAAGCGCA 2512
Db 2434 AGCTTACAGCAGCAGCTCAACGAGGAGCAACCGGTACCTCAAGTACCAACGCGCA 2493
Qy 2513 CGCGGATTTCAAGAGCGTGTGCAAGAAATACGCTTTTGGGGGCAACTCGGGGACAGC 2572
Db 2494 CGGAGATTTCAAGAGCGCTTAAAGAAATACGCTTTTGGGGGCAACTCGGGAGAGC 2553
Qy 2573 AGTCTTCAGAGCAAGAGCGGGTCTCGAACTCTGCTGCTGAGTGAAGAGCGCTAA 2632
Db 2554 AGTCTTCAGAGCAAGAGCGGGTCTTGAACCTCTGGGCTGAGTGAAGAGCGCTAA 2613
Qy 2633 GAGCGCTCTGGAAGAAAGTCTCGGTAGAGCACTGCGCAACAGGCAACTCTCTCTC 2692
Db 2614 GAGCGCTCTGGAAGAAAGTCTCGGTAGAGCACTCTCTGAGGCAACTCTCTCTC 2673
Qy 2693 GGGCATTCGAGAAAGAGCGGAGCGGCTTAAAGAAAGATCAATTTTGTGAGACTGG 2752
Db 2674 GGGAAACGGAAAGCGGGGCGAGGCTGCAAGAAAGATTAATTTTGTGAGACTGG 2733
Qy 2753 GCACTCAGAGTCACTGCTCCGATCCAACTCTGAGGAACTCTCAGCAACCCCGCTGC 2812
Db 2734 AGAGCGAGTCACTGAGTCTGAGCCCGAGCTCTGAGACGCAACGAGCAACCCCTCTGG 2793
Qy 2813 TGTGGGATCTAATCAATGCTTCAAGCGGTGAGGCACTAATGAGCAATTAAGAGG 2872
Db 2794 TGTGGGATCTAATCAATGCTTCAAGCGGTGAGGCACTAATGAGCAATTAAGAGG 2853
Qy 2873 CGCGGACGAGTGGGTATGCTCAGAGAAATTTGCAATTTTGTGAGTCTGAGGCGA 2932
Db 2854 CGCGGACGAGTGGGTATGCTCAGAGAAATTTGCAATTTTGTGAGTCTGAGGCGA 2913
Qy 2933 CAGAGTCAATCAACGACGAGCACTGAGGCTTGGCCACTTACAAATTAACAACCTCTA 2992
Db 2914 CAGAGTCAATCAACGAGCACTGAGGCTTGGCCACTTACAAATTAACAACCTCTA 2973

Qy 2993 CAAGCAATCTCCAGTCTTCAACGGGGGCGAGCAACGAACTACTTGGCTACAG 3052
Db 2974 CAACAAATTTCCAGCAATCA---GAGCTTCGAAAGCAATCACTACTTGGCTACAG 3030
Qy 3053 CACCCCTGGGGTATTTGATTTCAAGATTTCCAGTCTGCACTTTTCAACAGTACTG 3112
Db 3031 CACCCCTGGGGTATTTGATTTCAAGATTTCCAGTCTGCACTTTTCAACAGTACTG 3090
Qy 3113 GCAGGATCTCAACAAATTTGGGATTTCCGGCCCAAGAGACTAATCTCAACTCTT 3172
Db 3091 GCAGGATCTCAACAAATTTGGGATTTCCGGCCCAAGAGACTAATCTCAACTCTT 3150
Qy 3173 CAACATCAAGTCAAGAGGTCAAGCAGATGAGGCTCAACATCGCTAATACCT 3232
Db 3151 TACATTTCAAGTCAAGAGGTCAAGCAGATGAGGCTCAACATCGCTAATACCT 3210
Qy 3233 TACAGCAGGTTCAAGTCTTCTGAGCTCGAGTACAGCTTCCGTACGCTCGCTC 3292
Db 3211 TACAGCAGGTTCAAGTCTTCTGAGCTCGAGTACAGCTTCCGTACGCTCGCTC 3270
Qy 3293 TGGGACAGGGGTGCTCTCCCTCGTTCCCGGCGAGCTGTTATATTCGCAATACG 3352
Db 3271 GGGGATCAAGGATGCTCTCCCGCTTCCAGAGAGCTTCAATGATGCTCAAGTATG 3330
Qy 3353 CTACCTGAGGCTCAATGAGGAGGAGCGGAGGAGCTTCAATCTTTTACGCTGGA 3412
Db 3331 ATACCTGAGGCTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3390
Qy 3413 ATATTTCCCTTCTCAGATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3472
Db 3391 GATCTTCCCTTCTCAGATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3450
Qy 3473 GGAAGTCTCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3532
Db 3451 GGAAGTCTCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3510
Qy 3533 TCTCATGACCAATACCTGATTAACCTGAAAGAAATCAATCAATGAGGAGGAGGAG 3592
Db 3511 TCTCATGACCAATACCTGATTAACCTGAAAGAAATCAATCAATGAGGAGGAGGAG 3570
Qy 3593 AAACAGGAGTCTGCTGTTAGCGGTGGGCTCTCAGCTGAGTGTCTTCAAGCCCAAA 3652
Db 3571 GAGTCAAGGCTTCAATTTTCTCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3630
Qy 3653 CTGGCTACTGGAACCTGTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3712
Db 3631 CTGGCTACTGGAACCTGTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3690
Qy 3713 CAACAGCAATTTTACCTGAGCTGAGTCTCAAAATTAACCTCAATGAGGAGGAGG 3772
Db 3691 CAACAGCAATTTTACCTGAGCTGAGTCTCAAAATTAACCTCAATGAGGAGGAGG 3750
Qy 3773 CATCAACCTGGAACCTGTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3832
Db 3751 GGTGAATC-----CGGCATGAGCAAGCCCAAGAGGAGGAGGAGGAGGAGGAG 3804
Qy 3833 GAGCGGTGATGATTTTGGAAAGAGGCGCGGAGCTTCAACATGATGAGCA 3892
Db 3805 GAGCGGTGATGATTTTGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3864
Qy 3893 TGTCAATTTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3952
Db 3865 GGTCAATTTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3924
Qy 3953 TGGGACGTTGAGTCAATTTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4012
Db 3925 TGGTCTGATTTTCAACCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3984
Qy 4013 TGTATGAGGATTAATCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4072
Db 3985 CACACAGGCGTCTTCCAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4044

QY 4073 CATTTGGGCGCAAAATTCCTCAGACAGATGACACTTTGACCCGCTCTCTTATGAGCGG 4132
 Db 4045 CATCTGGGCAAAAGTTCACACGAGCGACATTTTCACTCCCTCTCCCTCAGGGG 4104
 QY 4133 CTTTGGACTCAAGAACCCGCTCTCAGATCTCTCATCAAAAACAGCGCTGTTCGGAA 4192
 Db 4105 ATTGGGACTTAAACACCTCTCCCAAGATTTCTCATCAAGAAACACCGGTAAGCGAA 4164
 QY 4193 TCCCTCCGGGAGTTTTCAGTCAAAAGTTGCTTCATCTCACTCCCAATCTCCACAG 4252
 Db 4165 TCTTTCACCACTTCAAGTGGGCGAAAGTTTGTCTTCTTCAACACAGTACTCCAGGG 4224
 QY 4253 ACA-AGTAGAGTGGAAATTGAATGAGAGCTGCAAGAAAGAAAGAGCAAGCGCTGGAATC 4311
 Db 4225 ACAAGGTAGAGGTGAGATGAGTGGAGCTGCAAGAAAGAAAGAGCAAGCGCTGGAATC 4284
 QY 4312 CGAAGTGCAGTACACATCCAAATTATGCAAAATCTGCCAAGTTGATTTTACTGTGACA 4371
 Db 4285 CCGAAATTCAGTACACTTCCAACTAACAAAGTCTGTTAATCTGGAATCTCGTGATA 4344
 QY 4372 ACAATGACTTTATAGAGCTGCGCCGATTTGACACCCGTTACTTACCCGCTCCCTGT 4431
 Db 4345 CTAAATGGGTATTCAGAGCTCGCCCATTTGGACACAGATACCTGATTAATCTGT 4404
 QY 4432 AATTACGTTTATCAATTAACCGGTTGATTCTTTAGTTGAATTTGTCTCTCTGTC 4491
 Db 4405 AATTGCTTTGTTAATCAATAAACCGTTTAATTCGTTTCACTTGAATTTGTCTCTGGTA 4464
 QY 4492 TTTCTTATCTTATCGGTTACCAATGTTATGCTTACATTAATCTGCTTGTGGCTTCG 4551
 Db 4465 TTTCTTCTTATCTTATGTTTCCATGCTTACATGATTAATGATACATGCGGGTTAATCAATA 4524
 QY 4552 CGATTAAGAATTAGATGAGTTTACCCCTAGATGAGTGGTCCCACTCCCTCTCTG 4611
 Db 4525 ACTACAAAGG-----ACCCCTAGATGAGTGGTCCCACTCCCTCTCTG 4568
 QY 4612 CGCGCTCGCTCGCTCGTGGTGGGCTCGCGACCAAAAGTCCCGACAGCGGACAGCTCTGCT 4671
 Db 4669 CGCGCTCGCTCGCTCGTCACTGAGGCGGGCGACCAAAAGTCCCGACAGCGGCTTGGCC 4628
 QY 4672 CTGCGGCGCCCAAGGAGCGAGCGGCGAGAGGAGGAGTGGCGAA 4718
 Db 4629 CGGCGGCGCTCAGTACGAGCGAGCGGCGGAGGAGGAGTGGCGAA 4675

RESULT 12
 AA2CG 4675 bp ss-DNA linear VRL 27-APR-1993
 LOCUS Adeno-associated virus 2, complete genome.
 DEFINITION J01901.M12405.M12468.M12469
 ACCESSION J01901.1 GI:209616
 VERSION 1
 KEYWORDS alternative splicing; complete genome; major coat protein.
 SOURCE Adeno-associated virus 2 DNA from human HeLa cells.
 ORGANISM Adeno-associated virus 2
 REFERENCE 1 (bases 4532 to 4675)
 AUTHORS Samulski, R.J., Srivastava, A., Berns, K.I., and Muzyczka, N.
 TITLE Rescue of adeno-associated virus from recombinant plasmids: gene
 JOURNAL Cell 33 (1), 135-143 (1983)
 MEDLINE 84282662
 PUBMED 6088052
 REFERENCE 2 (bases 1 to 4675)
 AUTHORS Srivastava, A., Lusby, E.W., and Berns, K.I.
 TITLE Nucleotide sequence and organization of the adeno-associated virus
 JOURNAL J. Virol. 45 (2), 555-564 (1983)
 MEDLINE 83164299
 PUBMED 6300419
 FEATURES
 SOURCE location/Qualifiers
 1..4675
 /organism="adeno-associated virus 2"
 /db_xref="taxon:10804"

repeat_region 1..145
 /note="5' inverted terminal repeat"
 misc_feature 42..83
 /note="flip oriented DNA"
 mRNA 287..4447
 /note="major coat protein A mRNA (alt.)"
 join(321..1906,2228..2252)
 CDS /note="major coat protein A' (alt.)"
 /codon_start=1
 /protein_id="AAA42372.1"
 /db_xref="GI:209617"
 /translation="MGFEYELIVKPSDLDGHLPGISDSFVNVAKEWELPDSMD
 LNLIEQAPLVAEKLORDPLETMRVSKAPALFPQFEKGSYFHMVLVETTVGKS
 MVGRFLSQREKLORIYRGLEPTLPNNFATVTKRAGGAGKAVDECTIPVLLPK
 TOPELQAMNTMEQVLSACINLTERRLVAOHLTVSOEQEKENQNPSPDAPVRS
 KTSARFMELVGMVLDKGIITSEKQIODEQASYISFNAAISRSQIQAALDNACKIMSL
 KTAPADLVGOQVEDISSNRIYKILENGYDQVYASVFLGATKFGKNTIWLFG
 PATTGKTNIAEALIAHTVPEYGVNMTNENPPENDCVDKNVIMWEGKMTAKVESAKA
 ILGGSFVRVDQCKSSAOIDPPEVITSNMCAVITDGSSTTEHQQLODRMFKEEL
 TRILDHPGKVTQOEYKDFPRMAKDIVEVEHEFYKKGAKKRPAPSDADISEPKRV
 RESVAQPSDSEASINVDRLARGHSL"
 321..2186
 /note="major coat protein A"
 /codon_start=1
 /protein_id="AAA42374.1"
 /db_xref="GI:209619"
 /translation="MPGEYELIVKPSDLDGHLPGISDSFVNVAKEWELPDSMD
 LNLIEQAPLVAEKLORDPLETMRVSKAPALFPQFEKGSYFHMVLVETTVGKS
 MVGRFLSQREKLORIYRGLEPTLPNNFATVTKRAGGAGKAVDECTIPVLLPK
 TOPELQAMNTMEQVLSACINLTERRLVAOHLTVSOEQEKENQNPSPDAPVRS
 KTSARFMELVGMVLDKGIITSEKQIODEQASYISFNAAISRSQIQAALDNACKIMSL
 KTAPADLVGOQVEDISSNRIYKILENGYDQVYASVFLGATKFGKNTIWLFG
 PATTGKTNIAEALIAHTVPEYGVNMTNENPPENDCVDKNVIMWEGKMTAKVESAKA
 ILGGSFVRVDQCKSSAOIDPPEVITSNMCAVITDGSSTTEHQQLODRMFKEEL
 TRILDHPGKVTQOEYKDFPRMAKDIVEVEHEFYKKGAKKRPAPSDADISEPKRV
 RESVAQPSDSEASINVDRLARGHSL"
 873..4447
 /note="major coat protein A mRNA (alt.)"
 join(993..1906,2228..2252)
 CDS /note="major coat protein Aa (alt.)"
 /codon_start=1
 /protein_id="AAA42373.1"
 /db_xref="GI:209618"
 /translation="MELVGMVLDKGIITSEKQIODEQASYISFNAAISRSQIQAAL
 NACKIMSLKTAPDLVGOQVEDISSNRIYKILENGYDQVYASVFLGATKFGKNTIWLFG
 PATTGKTNIAEALIAHTVPEYGVNMTNENPPENDCVDKNVIMWEGKMTAKVESAKA
 ILGGSFVRVDQCKSSAOIDPPEVITSNMCAVITDGSSTTEHQQLODRMFKEEL
 TRILDHPGKVTQOEYKDFPRMAKDIVEVEHEFYKKGAKKRPAPSDADISEPKRV
 RESVAQPSDSEASINVDRLARGHSL"
 993..2186
 /note="major coat protein A' (alt.)"
 /codon_start=1
 /protein_id="AAA42375.1"
 /db_xref="GI:209620"
 /translation="MELVGMVLDKGIITSEKQIODEQASYISFNAAISRSQIQAAL
 NACKIMSLKTAPDLVGOQVEDISSNRIYKILENGYDQVYASVFLGATKFGKNTIWLFG
 PATTGKTNIAEALIAHTVPEYGVNMTNENPPENDCVDKNVIMWEGKMTAKVESAKA
 ILGGSFVRVDQCKSSAOIDPPEVITSNMCAVITDGSSTTEHQQLODRMFKEEL
 TRILDHPGKVTQOEYKDFPRMAKDIVEVEHEFYKKGAKKRPAPSDADISEPKRV
 RESVAQPSDSEASINVDRLARGHSL"
 1853..4447
 /note="major coat protein B mRNA (alt.)"
 1907..2227
 /note="major coat protein A intron"
 2810..4324
 /note="major coat protein B"
 /codon_start=1
 /protein_id="AAA42376.1"
 /db_xref="GI:209621"
 /translation="MATGSGAPMADNNEGADGVGNSGNMHCSTYMGGRVITTSRTT

QY	1917	CCGACAGGTAACCAAAACAATTTCTGTCACGGGGAGATGCTTCAATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAATGTTCTGTCAGTGGGACATGATTTGATGCTGTTCCCT	1959
QY	1977	GCAACACATGCAAGAGAAATGAATCAGAATTTTCAACATTTGCTTCACGACGAGAG	2036
Db	1960	GCAGACATTCGAGAGAAATGAATCAGAATTTCAAAATATCTGCTTCACCTACGGACGAAAG	2019
QY	2037	ACTGTTCAAGATGCTTCCCGGCGTGTCAGAATTTCAACCGGT---CGTCAAAAGAGA	2093
Db	2020	ACTGTTTAGAGCTTCTCC---CGGTCAGAATCTCAACCCGTTTCTGTGTCAAAAGG	2076
QY	2094	CGTATCGGAACTCTGTGCAATTCATCTGCTGGGCGGGCTCCGCAATGCTTCTCT	2153
Db	2077	CGTATCAAGAACTGTGCTCAATTCATTCATATGGAAGAAAGTGCABA---CGTTGCA	2133
QY	2154	CGGCTCGCATCTGCTCAACGTGACCTGTGATGACTGTGTGTTCTGAGCAATTAATGACTT	2213
Db	2134	CTGCTCGCATCTGCTCAATGTGATTTGGATGATCGATCTTTGAAACAATTAATGATTT	2193
QY	2214	AAACAGGATAGGCTGCGGATGGTATCTTCCAGATGGCTGCSAGACAACTCTGTGAG	2273
Db	2194	AAATCAGGATAGGCTGCGGATGGTATCTTCCAGATGGCTGCSAGACAACTCTCTGAA	2253
QY	2274	GGCATTTCCGAGTGTGTGGGACTTTGAAACCTTGAGGCCGGAAGCCGAAAGCCAAACGACAA	2333
Db	2254	GGAAATTAAGCAGTGTGTGGAGCTCAAACTGTGCCACCAACCAACCAAGCCGACAGCGG	2313
QY	2334	AAGCAGAGACAGCGCCGGGCTTGTGCTTCTGTGCTACAGTACTCTGGACCTTTCAAC	2393
Db	2314	CATTAAGGACGACAGAGGGGCTTGTGTCTCTGTGGTACAGATACCTTGGAACCTTTCAAC	2373
QY	2394	GGACTCTGACAAAGGGGAGGCCCTCAACGCGGCGAGACGACGCGCCCTGAGACG-ACAA	2452
Db	2374	GGACTCTGACAAAGGAGAGGCCGCTTCAACAGAGAGAGCGCGGCCCTGAGAGACGTACA	2433
QY	2453	GGCTTACGACAGCAAGCTTCAAAAGCGGTTGACATTCGTACTGTGCGGTATTAACACACGCCGA	2512
Db	2434	AGCTTACGACCGCGAGCTCTGACAGCGGAGACAACTCGTACCTCAAGTCAACACAGCCGA	2493
QY	2513	CGCGAGTTTCAAGGAGCGCTGTGCAAGAAAGATGCGTCTTTTGGGGCAACCTCGGGCGAGC	2572
Db	2494	CGCGAGTTTCAAGGAGCGCTTAAAGAAAGATGCGTCTTTTGGGGCAACCTCGAGCGAGC	2553
QY	2573	AGTCTTCAAGGCAAGAAAGCGGTTCTCGAACCTTCGCTGTGTTGAGAAAGCGCTAA	2632
Db	2554	AGTCTTCAAGGCGAAAGAGGGTTCTTGAACTCTGGGCGTGTGTGAGAACTGTATA	2613
QY	2633	GACCGCTCTTGAAAAGAAACGTCCGCTTAGAGACAGTCCGACAAAGGCCAGACTCCCTCTC	2692
Db	2614	GACCGCTCCGGGAAAAAAGAGGCGCGTAGAGCACTCTCTGTGAGCCAGACTCTCTCTC	2673
QY	2693	GGGCACTCGGCAAGACAGGCGACAGCAACCCGCTTAAAAAGAGACTCAATTTTGGTCAGACTGG	2752
Db	2674	GGGAACCGGAAAGGCGGCGACAGCAACCTGTGCAAGAAAGATTTGGTCAGACTGG	2733
QY	2753	CGACTCAAGTCAATGCCCGATCCCAACCTTCTGGAGAACCTTCAGCAACCCCGCTGC	2812
Db	2734	AGACCCAGACTCAATGATCTGACCCCCAGCTCTCGGACAGCTCACAGCAGCCCTCTTGG	2793
QY	2813	TGTGGGACCTTACTCAATAGGCTTCAAGCGGTGGGCGACCAATGGCAGCAATTAAGAAAG	2872
Db	2794	TTTGGGAATTAATCAGATGCTCTACAGGCAGTGGCGACCAATGGCAGCAATTAAGAAAG	2855
QY	2873	CGCCGACGGAAGTGGGTAAATGCTCAGAAATTTGGATTGCGAATTCACATGGCTGGGCGA	2932
Db	2854	GGCCGACGGAAGTGGGTAAATTCCTCGGAAATTGGCATTTGCGATTGCAATGGATGGGCGA	2913
QY	2933	CAGAGTCATCAACCAACGACACCCGACCTGTGGGCTTGGCCCACTACATTAACACCTCTA	2992
Db	2914	CAGAGTCATCAACCAACGACACCCGAACTGTGGGCTTGGCCCACTACATTAACACCTCTA	2973

QY	2993	TAAGCAATCTCCAGTGGTTCAACGGGGGGCCAGAAAGACAACAATCTTCGGTCAAG	3092
Db	2974	CAAACTAATTTCCAGCCAAATCA---GGAAGCTCGAAGCAATATCAATCTTGGCTACAG	3030
QY	3053	CACCCCTGGGGGATTTTGAATTTCAACAGATTCACATGCGCACTTTTCACCACTGACTG	3112
Db	3091	CACCCCTTGGGGGATTTTGAATTTCAACAGATTCACATGCGCACTTTTCACCACTGACTG	3090
QY	3113	GCAGGACTCATCAACAATTTGGGGATTTCCGGCCCAAGACTCAACTTTCAAACTCTT	3172
Db	3091	GCAAGACTCATCAACAACAATTTGGGGATTTCCAGCCCAAGGACTCAACTTCAAGCTCTT	3150
QY	3173	CAACATCCAATCAAGAGGACTCAAGAGATGATGGCGTCAACAACATCGCTAATACT	3232
Db	3151	TAACTTTCAAGTCAAAAGAGTCAAGAGATGAGCGATACAGACGATTTGGCAATTAAGT	3210
QY	3233	TACCAAGCAGGTCAAGTCTTCTCGAAGCTGGAGTACCAAGTTCGTAAGTCTCGGCTC	3292
Db	3211	TACCAAGCAGGTCAAGGTGATTTTACGTACTGGATACAGTCCCGTAAGTCTCGGACTC	3270
QY	3293	TGGCAGCAGGGGCGCTCCCTCGGTTCCGGGGGAGAGTGTATGATTCGGCAATACGG	3352
Db	3271	GCGCATTAAGAGATGCTTCGCCGCTTCCAGCAAGATCTTCAATGGTGCACAGTATGG	3330
QY	3353	CTACCTGACGCTCAACAATGGCAGCAGCAGCCGTGGAGCTTCATCTTTTATGCTCTGA	3412
Db	3331	ATACCTCACTCTGAACAACGGGAGTCAAGGCGATGAGAGCTCTTCATTTTATCTCTGA	3390
QY	3413	ATATTTCCCTTCTCAGATGCTGAGAAAGGGCAACAATTTTACCTTCAGCTACACTTGA	3472
Db	3391	GTACTTCTCTTCTCAGATGCTGCGCTACCCGAAACAATTTTACCTTCAGCTACACTTGA	3450
QY	3473	GGAAGTGCCTTCCACAGAGCTACCGGCAAGCAGCAGAGCTTGAGCCGGCTGAATGCC	3532
Db	3451	GGAAGTTCCTTCCACAGAGCTACCGCTCAAGCAGAGCTTGAGCCGGCTGAATGCC	3510
QY	3533	TCTCATGACCAATTAAGTATTAATCTGGAACAGAACTCAAAATAGTCCGAAGTGGCCA	3592
Db	3511	TCTCATGACCAAGTACCTGATTAATCTTGAAGCAAACTCAAGTGAACACACAC	3570
QY	3593	AAACAAGGACTTGTGATTTAGCCGTGGGTCTCCAGCTGGGCAATGCTGTGACGCCAAAA	3652
Db	3571	GCAAGTCAAGGTTTAGTATTTCTCAGGCGCGAGCGAGTGAATTTGGGACAGTCTTAGAA	3630
QY	3653	CTGGCTACTGAGACTGTTATATGCGCAGACAGCGCGTTCTTAAACAAAAACAGACAACA	3712
Db	3631	CTGGCTTCTGAGACCTGTTATCTCCCGCAGCGAGTATCAAAAGCAATCTGGGATTAACA	3690
QY	3713	CAACAGCAATTTTACTGGAATGGTGCTTCAAAATATACTCAATGGGCGTGAAATCAT	3772
Db	3691	CAACAGTGAATATCTGAGACTGAGGCTAACCAATACCTCAATGGCAGAGACTCTCT	3750
QY	3773	CATCAACCTTGGAAGTGTATGGCTTCAACAAGACGAGCAGACAAAGTCTTTCCCAT	3832
Db	3751	GGTGAATC-----CGGCAATGGCAAGCACAAAGACGATGAAGAAAGTTTTTCTCA	3804
QY	3833	GAGCGGTATCATGATTTTGGAAAAAGAGCGCGGAGCTTCAAACTAGATTTGGACAA	3892
Db	3805	GAGCGGGTTCATCTCTTGGGAAGCAAGGCTCAGAGAAAACAATATGTGAACATTTGAAA	3864
QY	3893	TGTCAATGATTCACAGCAAGAGAAATTTAAAGCACTTAACCTGTGGCAACGAAATTT	3952
Db	3865	GGTATGATTTACAGCAAGAGAAATTCGAAACAACAATTCGCTGGCTACAGAGAGTA	3924
QY	3953	TGGGACGTTGCAAGTCAATTTCCAGAGACGACACAGACCTTCGACCCGAGATGTGCA	4012
Db	3925	TGTTTCTGTATCTCAACACTCCAGAGAGGCAACAGCAAGCAAGCTTACCGCAGATGTCAA	3984
QY	4013	TGCTATGGGAACATTAAGTGGCAAGTGTGGCAAGATAGAGAGGTAACTCGAGAGGGTCC	4072
Db	3985	CACACAAAGGCTTCTTCCAGGCAAGTGTGGCAAGACAGAAATGTATACCTTCCAGGGGCC	4044
QY	4073	CATTTGGGCCAAATTTCTCAACAGATGACATTTTCAACCCGCTCTCTTATGGGCGG	4132

Db 4045 CATCTGGGCAAAAGTTCACACAGCAGGACATTTTACCCCTCTCCCTCATGSGTGG 4104
Qy 4133 CTTTGACCTAAGAACCCGCTCTCTCATGATCTCTATCAAAAACAGCGCTGTCTCTGGAA 4192
Db 4105 ATTGCGACTTAAACACCCCTCTCCACAGATTCATCAAGAACACCCGGGTACTTGGAA 4164
Qy 4193 TCTCCGCGGAGTTTTCAGTCAAAAGTTTGTCTTCAATCATCAACCAATACTCTCAAGG 4252
Db 4165 TCTTTCGACCACTTCAAGTGGCGGAAAGTTTGTCTTCTTATCAACAGTACTCTCAAGG 4224
Qy 4253 ACA-AGTGAAGTGAAGTGAATGAATGAGCTGCAGAAAGAAACAGACAGCGCTGAATC 4311
Db 4225 ACACGCGTCAAGTGAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAATC 4284
Qy 4312 CCGAAGTGAAGTGAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4371
Db 4285 CCGAAGTGAAGTGAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4344
Qy 4372 ACATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4431
Db 4345 CTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4404
Qy 4432 AATTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4491
Db 4405 AATTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4464
Qy 4492 TTCTTATCTTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4551
Db 4465 TTCTTATCTTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4524
Qy 4552 CGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4611
Db 4525 ACTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4568
Qy 4612 CGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4671
Db 4569 CGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4628
Qy 4672 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4718
Db 4629 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4675

RESULT 13
AX205073 LOCUS AX205073 7557 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 3 from Patent WO0155361.
ACCESSION AX205073
VERSION AX205073.1 GI:15394314
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 7557)
AUTHORS Hardy, S. F.
TITLE Recombinant aav packaging systems
JOURNAL Patent: WO 015361-A 3 02-AUG-2001;
Chiron Corporation (US)
FEATURES
source 1..7557
/organism="unidentified"
/db_xref="taxon:32644"
/note="recombinant DNA"
BASE COUNT 1970 a 1940 c 1953 g 1694 t
ORIGIN
Query Match 63.4%; Score 2993.4; DB 6; Length 7557;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 3633; Conservative 0; Mismatches 841; Indels 45; Gaps 9;
Qy 18 GCGGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 77
|||||
Db 18 GCGGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 77
|||||

Db 4 GCGGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 63
Qy 78 TCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 137
Db 64 TCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 123
Qy 138 GGGTAAT-----CGCGAAGCGCTCCACGCTCCCGCTC 172
Db 124 TGGGACTGTGTACATTAATCTGTATATATATATATATATATATATATATATATATATAT 183
Qy 173 AGCGCTGAGTAAATTAATGATAGG---GAGTGTCTGTATATATATATATATATATATAT 229
Db 184 AGTGTGAGCTGATTAATGATAGG---GAGTGTCTGTATATATATATATATATATATATAT 243
Qy 230 GCTTTGCGGCAATTTTGGCAGACCAAGTGGCCATTAAGGTATATATATATATATATATATAT 289
Db 244 G-TTTTGGCAGATTTTGGCAGACCAAGTGGCCATTAAGGTATATATATATATATATATATAT 302
Qy 290 AGCAGATCTCATTTTGG---ACCGGAAATTTGAAAGCAGCAGCAGCAGCAGCAGCAGCAG 348
Db 303 CGCAGGAGTCTCATTTTGAAGCGGAGGTTTGAACCGGAGCGGCGCATGCGGGGTTTGA 362
Qy 349 CGAGATCTGATCAAGTGTCCGAGCGACCTGAGCAGACCTTCCGCGCATTTTCTGATCTC 408
Db 363 CGAGATTTGATTAAGTGTCCCGAGCGACCTTGAACGAGCATCTCCGCGCATTTTCTGACAG 422
Qy 409 GTTTGTAGCTGT 468
Db 423 GTTTGTAGCTGT 482
Qy 469 GAATCTGATTTGAACAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 528
Db 483 GAATCTGATTTGAACAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 542
Qy 529 CCAATGCGCGCGGTGATAGGCGCCCGAGAGCGCTCTTTCTTTGTTGATTTGAGAAAGG 568
Db 543 GGAATGCGCGCGGTGATAGGCGCCCGAGAGCGCTCTTTCTTTGTTGATTTGAGAAAGG 602
Qy 589 CGAGTCTTACTTCCACTCTCATTTTGTGTGAGACACAGGGGTCAATTCATGTGTCT 648
Db 603 AGAGACTACTTCCACTCTCATTTTGTGTGAGACACAGGGGTCAATTCATGTGTCT 662
Qy 649 GGGCGGCTTCTGAGTCAAGTGAAGCAAGTGTGTGAGACACATCTTACCGCGGATCGA 708
Db 663 GGGAGCTTCTGAGTCAAGTGAAGCAAGTGTGTGAGACACATCTTACCGCGGATCGA 722
Qy 709 GCGGACCTTCCCAACTGTTTGGCGGTGACCAAGACCGGTATGTGCGCGGAGGAGAA 768
Db 723 GCGGACCTTCCCAACTGTTTGGCGGTGACCAAGACCGGTATGTGCGCGGAGGAGAA 782
Qy 769 CAAGGTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGGAGCT 828
Db 783 CAAGGTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGGAGCT 842
Qy 829 GCAAGTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGGAGCT 888
Db 843 CAGTGTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGGAGCT 902
Qy 889 CAAGCGCTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGGAG 948
Db 903 TAAACGTTGTGTGTGAGACGATGTCAATCTCCCAATCTCTCTGCGGCAAGATCAAGCGG 962
Qy 949 GAATCTGAAACCCCAATTTTGAAGCGGCTGTATCTCGGTCAAAAACCTTCCGCGGCTAT 1008
Db 963 GAATCTGAAATCCCAATTTTGAAGCGGCTGTATCTCGGTCAAAAACCTTCCGCGGCTAT 1022
Qy 1009 GAGAGT 1068
Db 1023 GAGAGT 1082
Qy 1069 GAGAGT 1128
Db 1083 GAGAGT 1142


```

Db      3291 TCGGCTGGCGCATCAAGAGATGCTCCCGCGTTCCAGACAGAGTCTTATGGTGCAC 3350
Qy      3346 AATAAGGCTACCTGACCGCTCAAGAAATGAGCGCAAGCGGTGGAGCTTCACTCTTACT 3405
Db      3351 AGTATGAGTACCTACCTGAAACAGGAGTCAAGGAGTGAAGCGCTCTTCACTTTACT 3410
Qy      3406 GCCTGAAATATTTCTCTCTCAGATGCTGAGAACGGGCAACAATTACTTCACTGACTA 3465
Db      3411 GCCTGAGATCTTCTCTCTCAGATGCTGAGAACGGGCAACAATTACTTCACTGACTA 3470
Qy      3466 CTTTGAAGAAATGCTCTTTCACAGACTACGCGCAACGCGAGCTTGGACCGGCTGA 3525
Db      3471 CTTTGAAGAAATGCTCTTTCACAGACTACGCGCAACGCGAGCTTGGACCGGCTGA 3530
Qy      3526 TGAATCTCTCACTGACCAATATCTGATTAATCTGAAACAAATCAATCAATGCTG 3585
Db      3531 TGAATCTCTCACTGACCAATATCTGATTAATCTGAAACAAATCAATCAATGCTG 3590
Qy      3586 GTGCGCCAAAACAAGGACTGCTGTTAGCGGTGCTCCAGCTGGCATGCTGTTGAGC 3645
Db      3591 CCACACGAGCTCAAGGCTTCAAGTTTCTAGCGCGGAGAGTGAATTCGGACCACT 3650
Qy      3646 CCAAAAATGCTGCTACCTGACCTCTGTTATCGGACGAGCGGCTTCAAAAACAAA 3705
Db      3651 CTAGAAATGCTGCTTCTGACCTGTTACGCGCAGCAGCGATCAAGACATCTGCGG 3710
Qy      3706 ACACAAACAACAGCAATTTTACCTGACCTGTTGTTAAATATTAATCAATCTGAG 3765
Db      3711 ATACAAACAACAGCAATTTTACCTGACCTGTTGTTAAATATTAATCAATCTGAG 3770
Qy      3766 AATTCATCATCAACCTGACCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3825
Db      3771 ACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3830
Qy      3826 TTCCCATGAGCGGTGCTGATGATTTTGGAAAAGAGCGCGGAGCTTCAACACT 3885
Db      3831 TTCCCATGAGCGGTGCTGATGATTTTGGAAAAGAGCGCGGAGCTTCAACACT 3890
Qy      3886 TGAACATATGATGATTAACAGACGAAAGAAATTAAGCACTTACCTGCTGCTG 3945
Db      3891 TTAACAAAGTATGATTAACAGACGAAAGAAATTAAGCACTTACCTGCTGCTG 3950
Qy      3946 AAGATTTGGGACCGTGGAGTCAATTTCCAGACGACGACGACGACGACGACG 4005
Db      3951 AGCAGATATGTTCTGATCTTCAACCACTCCAGAGAGGCAACAGACGACGACG 4010
Qy      4006 ATGTGATGCTATGAGGATTAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCT 4065
Db      4011 ATGTGATGCTATGAGGATTAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCT 4070
Qy      4066 AGGATCCCATTTGGGCAAAATTTCTCAACAGATGACCTTCAACCTGCTGCT 4125
Db      4071 AGGATCCCATTTGGGCAAAATTTCTCAACAGATGACCTTCAACCTGCTGCT 4130
Qy      4126 TGGGCGGCTTTGAGCTCAAGAACCGGCTCTCTGATCTCATCAAAACACGCTGCT 4185
Db      4131 TGGGCGGCTTTGAGCTCAAGAACCGGCTCTCTGATCTCATCAAAACACGCTGCT 4190
Qy      4186 CTGCGAATCTCCGCGAGTTTCACTCAAAATTTGCTTCAATTCATCAACCAAT 4245
Db      4191 CTGCGAATCTCCGCGAGTTTCACTCAAAATTTGCTTCAATTCATCAACCAAT 4250
Qy      4246 CCACAGACAGATGATGAGTGAATTAATGAGGAGCTGCAAGAAAGAAACAGAG 4305
Db      4251 CCACAGACAGATGATGAGTGAATTAATGAGGAGCTGCAAGAAAGAAACAGAG 4310
Qy      4306 GGAATCCCGAAGTCAATCAATCAATTAATGCAAAATCTGCAACGTTGATTTACT 4365
Db      4311 GGAATCCCGAAGTCAATCAATCAATTAATGCAAAATCTGCAACGTTGATTTACT 4370
Qy      4366 TGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4425

```

```

Db      4371 TGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4430
Qy      4426 CCCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4485
Db      4431 ATCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4490
Qy      4486 CTGCTCTCTTATCTTATC 4504
Db      4491 TGGGATTTCTTCTTATC 4509

RESULT 14
AF369963
LOCUS       7327 bp    DNA    circular SYN 22-JUL-2001
DEFINITION Cloning vector pAAV-RC, complete sequence.
ACCESSION  AF369963
VERSION     AF369963.1  GI:14994089
KEYWORDS   Cloning vector pAAV-RC.
SOURCE     Cloning vector pAAV-RC.
ORGANISM   artificial sequences; vectors.
REFERENCE  1 (bases 1 to 7327)
AUTHORS   Grafsky, A.J. III.
TITLE     pAAV-RC for AAV Helper-Free System
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 7327)
AUTHORS   Grafsky, A.J. III.
TITLE     Direct Submission
JOURNAL   Submitted (16-APR-2001) Technical Services, Stratagene, 11011 N.
          Torrey Pines Rd., La Jolla, CA 92037, USA
          Location/Qualifiers
            1..7327
              /organism="Cloning vector pAAV-RC"
              /db_xref="taxon:161364"
              /lab_host="Escherichia coli"
              /note="used for recombinant DNA processing"
              131..1996
                /note="AAV-2 rep"
                /codon_start=1
                /transl_table=11
                /product="adenovirus-associated virus 2 rep protein"
                /protein_id="AAK76418.1"
                /db_xref="GI:14994090"
                /translation="MGPFEIIVIKVPSDIDELHPIGTSDFVNMVAKEMLPDSMD
                LNIIDAPLIVAEKLDRLITRMRSKAPDLFQPFKBSYFPMHVLVETGYS
                MUGRLSITREKLIQITRGIEPTLPNFAVTKRNGGKRVDECTIPEYLPK
                TOPELQWANTMEQYLSACLITERRKLVASHLTVSQIOBKENQNPNSAPVRS
                KTSARYMELVGMVLDKGTSEKOWIOEDQASYISFMAANSRSQIKALDNGKIMSL
                TKIAPDYLVGOOPVEDISSNRKIYKILNGDYDQYASVFLGATKFGKRTIMFG
                PATGKTNIABAIATHTVPYGCVMNTNENPPNDICVDMVIMMERGMTAKVYESAKA
                IIGSKVRVDOKKSSAQIDPPIVITSNTNCVAVDGNSTTFEHOPIODNRKREL
                TRLDHDFKVTKEQVDFRNKDHVVEHEFVYKKGAKRPPASDADISPTFRV
                RSVAPQSTSDAEASINVADRYONKSRVGNMLMFPQOCERNANQNSNICEPTROK
                DCLECPVSESPVSVKAYOKLCYIHHIMKVPACTACDLVNVLDLDCIFEQ"
                2013..4420
                  /note="cap VP1"
                  /codon_start=1
                  /transl_table=11
                  /product="capsid protein VP1"
                  /protein_id="AAK76419.1"
                  /db_xref="GI:14994091"
                  /translation="MAADGVLPDWDLEDTLSEGRIMWKLPQPPPKPARBHKDPSRG
                  IYLRGKYGLGPRNGDKEPVPVADDAALIEBHKAYRQIDSGDNPLTKNHDABEOR
                  RLKEDTSFGNIGRAVFOAKKRVLEPQGLVVEEVPKAPGKRPEVSPVBPSSGSG
                  KQGGQPARRLNFGQGDADSVDDPQGLPPAPAGLPTNTMATSGAPMDNNGA
                  DVGNSGSMWHDSTWMDGVITTSRTALPTYNHLTKYQISSGAGANDHYFEYS
                  TWGYPDENRPFHCFSPDMORLINNMGPEPLNFKLFINIOVKTEKONDGTTIAN
                  NLTSTQVPTDEYOLPYVLGASHQGLPPPADVNVVQYGLITLNNSSOAVGRSEF
                  YCLVPRSQMLRTGNNPTSTEDVPFSSVAGVLCITLNNPLIDVLYLSTNT
                  PSSTTQSLQSGASIDIRPQSRMLGVCITRQQRVSKTADANNNSYMTGATKY
                  HLNGRSLVNPAPAMASHKDEKFFPGSGVLPFGQSGEKTNVDIEKMTIDEEIR
                  TTNPAVTEYGSVSTNLGRNQAATADVNTQGLVPGMWQDRDVLQPIWAKIPIHT
                  DGHFHSPLMGFGILKHPPIILIKNTIPVAPNPSTTFSSAKPASFITYSTGQVSEI

```

```

rep_origin
CDS
EWEI,OKENSKRWNP,EI,QTTSYNNKS,VAND,FV,VDNNGVYSEBR,PIGTGYL,LRNL",
4705..5161
/note="fl origin"
5292..6152
/note="confers ampicillin resistance"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AAK76420.1"
/db_xref="GI:14994092"
/translation="MSIQHFRVALIPFPAACLPVFAHPETLVYKYKAQEDQLGARVGY
I,ELDNLSDGIESFPRRPEPRFPMWSGTFFKVLLCGAVLSIRIDACQEQ,IGRRIRH,SOND,VE
YSP,PEVKHI,TDMT,VI,RELCSA,I,ITMSD,NTAANLL,ITIGKREKLTAP,LAHNNQ,AVTL
DMWEEL,ELI,PDND,DTT,MPVAA,MT,LRKLT,TEL,LT,LA,SRQOL,ID,MEW,ED,KVAGCL
LRSALP,DAGGFI,INDKSG,AGERSG,RI,IALGPD,KRSR,IV,VI,TTGSAQ,VD,MRNR,QLA
ETGASAL,ITGHM"
6161..7019
rep_origin

```

	/note="COLE1 origin"
BASE COUNT	1903 a 1887 c 1912 g 1625 t
ORIGIN	

Query Match	61.0%;	Score 2878.4;	DB 12;	Length 7327;
Best Local Similarity	81.3%;	Pred. No. 0;		
Matches 3403; Conservative	0;	Mismatches 766;	Indels 15;	Gaps 5;

QY	324	GAGCAGCAGCATTGCGGGCTTCTACGAGATGCTGATCAAGGTGCGAGCGACTTGACG	383
Db	120	GGCGAGCGCCATGCGGGGTTTTCAGAGATTGTGATTAAAGGTCCCGACGCACTTGGACG	179
QY	384	AGCATCTGCCGGGCAATTTCTGACTCGTTGTGTAGCTGGTGGCCGAGAAAGGAAATGGGAGC	443
Db	180	AGCATCTGCCGGCAATTTCTGACACTTTGTGAACCTGGGTGGCCGAGAAAGGAAATGGGAGT	239
QY	444	TGCCCGCGGATTTCTGACATGATCTGAATCTGATTTAGACAGGCAACCCCTGACCTGTGGCG	503
Db	240	TGCCCGCCAGATTTCTGACATGATCTGAATCTGATTTGAGAGGCAACCCCTGACCTGTGGCG	299
QY	504	AGAACTGCAGGCGCATTTCTGTGTCCAAATGGCCGCGGTGAGTAAAGCCCGGAGGCC	563
Db	300	AGAACTGCAGGCGCATTTCTGACGGAATGGCCGCGGTGAGTAAAGCCCGGAGGCC	359
QY	564	TCTTCTTGTCAGTTTGAGAAAGGGCGAGTCCCTACCTTCACCTCCATATTTCTGTGGAGAA	623
Db	360	TTTCTTTGTGCAATTTGAAAGGAGAGAGTACTTCCACATGCACTGCACTGCTGTGTAATA	419
QY	624	CCACGGGGGTCAAAATCATTGTGTCTGGGCCCTTCTCTGATGATTAAGGACAAGCTGG	683
Db	420	CCACGGGGGTGAATTCATGTGTTTTGGGACGTTTTCTGTGATCAGATTCGCGAAAATCGA	479
QY	684	TGCAGACATCTACCGCGGGGATTCGAGCCGCACTTCGCCCACTGTGTTCCGCGTGAACCAAGA	743
Db	480	TTTCAGAGAAATTTACCGCGGGATTCGAGCCGCACTTTCGCAAACTGTGTTCCGCGTCACAAAAGA	539
QY	744	CGCGTAATGGCGCCGAGGGGGGAAACAAGTGTGTGAAGAGTGCTACATCCCACTAAC	803
Db	540	CCAGAAATGGCGCCGAGGGCGGAAACAAGTGTGTGAATGAGTGTCTACATCCCACTTAAT	599
QY	804	TCTGCGCCAGAACTCAGCCCGGACTGCAATGGGCGGTGAATTAATGATGAAGAGTATATAA	863
Db	600	TGCTCCCAAAAACCAAGCTGAGTCTCAATGGGCGGTGACTAATATGGAACAGTATTTAA	659
QY	864	GCGCTGTGTTGAATCTGGCCGAGCGCAAAAGCGCTGTGTGGCGGAGCAACTGATCCCACTCA	923
Db	660	GCGCTGTGTTGAATCTCAGGAGCGTAAACGGTGTGTGGCGGAGCATCTGAGGACGTGT	719
QY	924	GCAGAACCCAGAGAGAAACAAGAGAAATCTGAACCCCAATCTGACGGCGCTGTATCC	983
Db	720	CGCAGACCGAGAGAGAAACAAGAGAAATCCCAATCTGTATGGCGCGGTGATCA	779
QY	984	GGTCAAAAACCTCGCGCGCTACATGAGCTGTGTGGGTGGCTGTGTGAACCGGGGCATCA	1043
Db	780	GATCAAAAACCTTCAGCCAGAGTCAATGAGCTGTGTGGGTGGCTGTGTGAACAAGGGGATTA	839

QY	1044	CTCTCCGAGGAGCGATGGATCCAGAGAGACCGAGCGCTGTACATCTCTTCAACGCGCCTT	1103
Dp	840	CTTCGGAGGAAGCAATGATCCAGAGAGACCGAGCGCTGTACATCTCTTCAACGCGCCT	899
QY	1104	CCAACTCGCGGTCGCCAGATCAAGCGCGCTGTGCAACATGCGCGCAAGATCATGCGCTGA	1163
Dp	900	CCAACTCGCGGTCGCCAAATCAAGGCTGCTTGGACAAATGCGGGAAGATTAATGAGCTGGA	959
QY	1164	CCAAATCCGCGCCGATACCTAGGTAGCCCCCGCTCCGCCCGGGAACATTAAACCAACC	1223
Dp	960	CTAAACCCGCCCCGATACCTAGGTAGCCCCCGCTCCGCCCGGGAACATTAAACCAACATC	1019
QY	1224	GCATCTACCGCATCTGCTGAAACGGCTACGAACTGCGCTACGCGCGCTCGCTCTTTC	1283
Dp	1020	GGATTTATAAATTTTGGACTAAACGGGATACATCCCAATATGCGGCTTCCGCTTTTC	1079
QY	1284	TCGGCTGGGCGCCAGAAAAGGTTGCGGAAAGCCAAACACCATCTGCTGTTTGGCGCGCA	1343
Dp	1080	TGGGATGGGCGCCAGAAAAGTTGCGGCAAGAGAAACACCATCTGCTGTTTGGCGCTCAAA	1139
QY	1344	CCACGGGCAAGACCAATGCGCGAAGCCATTCGCCACGCGCTGCTTCTTACGCGCTGCG	1403
Dp	1140	CTACCGGAAAGACCAATGCGCGAAGCCATTCGCCACGCTTCTTACGCGGCTGCG	1199
QY	1404	TCAACTGGAACATGAAATTTTCCCTCAATGATTTGCTGACCAATGATGATCTGCT	1463
Dp	1200	TAAATGGAACCATGAAATTTTCCCTTCAACGACTGTGACAGATGATGATCTGCT	1259
QY	1464	GGGAGGAGGCGCAAGATACGCGCAAGGTCGTGAGTCCGCAAGGCCATTTCTCGCGCGCA	1523
Dp	1260	GGGAGGAGGGAAGATGACCGCCCAAGGTCGTGAGTCCGCAAGGCATTTCTCGAGAA	1319
QY	1524	GCAAGTGTCGCTGTGACCAAAAAGTGCATGTCGCCCAAGATGCAACCCACCCCGCTGA	1583
Dp	1320	GCAAGTGTCGCTGTGACCAAAAAGTGCATGTCGCCCAAGATGCAACCCACCTCCGCTGA	1379
QY	1584	TTCGTCACCTTCCAAACCAACATGTGCGCGGATGTGACGGAACAGACCACTTCTGAGC	1643
Dp	1380	TGTCACCTTCCAAACCAACATGTGCGCGCTGATGTGACGGAACCTCAACGACTTCTGAAAC	1439
QY	1644	ACCAGCAGCGCTGTGCAAGACCGGATGTTCAAAATTTGAACTCACCCGCGCTGTGAGCATG	1703
Dp	1440	ACCAGCAGCGCTGTGCAAGACCGGATGTTCAAAATTTGAACTCACCCGCGCTGTGATCATG	1499
QY	1704	ACTTTGGCAAGGTGACAAAGCAGGAATCAAGAATCTTTCGCTGGGCGCAGATCAAG	1763
Dp	1500	ACTTTGGGAAAGTGCACCAAGCAGGAATCAAGAATCTTTCGCGTGGCAAAAGGATACG	1559
QY	1764	TGACCGAGGTGGGCGATAGATGTTCTACGTCAGAAAGGTGAGGCCAAMAAAGACCGCGCC	1823
Dp	1560	TGTTTGGAGTGGACATGAAATTTCTACGTCAGAAAGGTGAGGCCAAMAAAGACCGCGCC	1619
QY	1824	CCGATGACGCGGSAATAAAGCGAGCCCAAGCGGCGCTGCCCTCATGCGCGATCCATCGA	1883
Dp	1620	CCGATGACGCGGATTAAGTGAAGCCCAAGCGGCGCTGCCCTCATGCGCGATCCATCGA	1679
QY	1884	CGTCAGACGCGGAAAGAGCTCCGCTGGAATTTGCGCAAGGTACAAAACAAATGTTCTC	1943
Dp	1680	CGTCAGACGCGGGA---AGCTTCGATCAACTACGCAAGCAAGTACCAAAACAAATGTTCTC	1736
QY	1944	GTCACGGGGGCAATCTCAATGCTGTGTTCCCTGCAAGACATGAGAGAGATGAATCAGA	2003
Dp	1737	GTCACGGGGGCAATCTCAATGCTGTGTTCCCTGCAAGACATGAGAGAGATGAATCAGA	1796
QY	2004	ATTTCACATTTTGCTTACGCAAGGACGAGAGCTGTTGAGATGCTTCCCGCGCGTGT	2053
Dp	1797	ATTTCACATTTTGCTTACCTACGCAAGAAAGCTGTTAGAGTGTCTTCC---CGTGT	1853
QY	2064	CAGAAATCAACCGGCT---CGTCAGAAAGAGAGCTATCGGAACTCTGTGCCATTCATC	2120
Dp	1854	CAGAAATCAACCGGTTTCTGTGTGTCAAAAAGGCGTATCAGAAACTGTGTACATTCATC	1913
QY	2121	ATCTGTGGGGCGGCGCTCCGGAATGCTTGTCTGCGCTCGATCTGTCAACCTGGAC	2180

Db	1914	ATATCATGGGAAAGGTGCCAAG--CGCTTGCACTGCTCGGATCTGGTCAATGTGCAATT	1970
Oy	2181	TGATGACTGTGTCTAGACCAATTAATGACTTAAACCAAGGTATGTGCTCCGATGTAT	2240
Db	1971	TGATGACTGTGCACTTTTGAAACAATAATGATTAAATACAGGTATGTGCTCCGATGTAT	2030
Oy	2241	CTTCCAGATTGGCTGTGAGAGCAACTCTCTCTGAGGGCAATTCGCAGTGTGGACTTGAA	2300
Db	2031	CTTCCAGATTGGCTGTGAGAGCACTCTCTCTGAGGAATTAACAGTGTGTGAACTCAAA	2090
Oy	2301	CCTGAGGCCCCGGAAGCCCAAGGCCAACGAGGAAAGCGAGCGACGCGCGGGGTCTGGTG	2360
Db	2091	CCTGAGCCCAACCAACCAAGGCCCCGAGGGGCAATAAGACGACGAGAGGGGTCTGTGTG	2150
Oy	2361	CTTCTGGCTCAAAAGTAACTTCGGAACCTTTCAAACGACTCGAACAGGGGGAACCCGTCAC	2420
Db	2151	CTTCTGGGTAAAGTAACTTCGGAACTCTTCAAACGGAACCTCGAACAGGAGAGACCGGTCAAC	2210
Oy	2421	GCGGCGGACGCAAGCGGGCCCTTGAGCAACAAGGCTCTAGACACAGACCTCAAAAGCGGT	2480
Db	2211	GAGGAGAGACGGCGCGCCCTCGAGACACAAGAGCTTCGACCGGCAAGCTCGAAGCGGGA	2270
Oy	2481	GACAACTCGGTACTGCGGTATATACAGCGCCGACCGCGAGTTTCAAGAGCGGTCCCAAGA	2540
Db	2271	GACAACTCGGTACTGCGGTATATACAGCGCCGACCGCGAGTTTCAAGAGCGGTCCCAAGA	2330
Oy	2541	GATACGTCTTTGGGGGCAACCTCGGGGAGACAGTCTTCAAGGCAAGAAACGGGTCTTC	2600
Db	2331	GATACGTCTTTGGGGGCAACCTCGGAAGAGAGTCTTCAAGGCAAGAAAGGGTCTTT	2390
Oy	2601	GAACTCTTCGGTCTGTGTGAGAGAGGCGCTAAAGAGCGTCTTGAAAGAAAGTCGGSTA	2660
Db	2391	GAACTCTTCGGGCTGTGTGTGAGGAACTCTTAAAGAGCGCTCCGGGAAAAAAGAGCCGSTA	2450
Oy	2661	GAGCAGTGGCCAAAGAGCCGAGACTCCCTCTCGGGCAATCGGCAAGACAGGCGACAGCC	2720
Db	2451	GAGCACTCTCTGTGGAGCCAGACTCTCTCTCGGGAAACGGAAAGGCGGGGCAACAGCTT	2510
Oy	2721	GCTAAAAAGAGACTCAATTTTGTGTGACAGTGGCGACTGAGAGTCAAGTCCCGATTCACAA	2780
Db	2511	GCAAGAAAAAAGATTGAAATTTTGTGTGACTGAGAGCGAGACTCAAGTACTGACCCCGAC	2570
Oy	2781	CCTCTCGGAGAACTCTCGAGAACCCCGGCTGTGTGGACCTTACTCAATAGGCTTCAGGC	2840
Db	2571	CCTCTCGGAGACCCACAGACAGCCCTCTGTGTGTGGGAACTAATGATGGCTTACAGGC	2630
Oy	2841	GGTGGCGGACCAATGGGCAACAATAACGAAGCGCCGACGGAGTGGGTAAATGCTTCAGGA	2900
Db	2631	AGTGGCGGACCAATGGGCAACAATAACGAAGGCGCCGACGGAGTGGGTAAATTCCTCGGGA	2690
Oy	2901	AATTGGCAATTCGATTCCACATGGGTGGGGCGACAGAGTCAATCACACGAGACCCGACCC	2960
Db	2691	AATTGGCAATTCGATTCCACATGGAGTGGGCGACAGAGTCAATCACACGAGACCCGAAAC	2750
Oy	2961	TGGGCTTTGGCCACCTTACAAATPACCACTCTPACAGCAAAATCTCAAGTCTTCAACGGGG	3020
Db	2751	TGGGCGCTGGCCACCTTACAAACCACTCTTACAAACAAATTTCCAGCCANTCA--GGA	2807
Oy	3021	GCCAGCAACGACAACTACTTCTTGGGCTTACAGCACCCCTTGGGGGTATTTTGATTTCAAC	3080
Db	2808	GCTTCGAACGACAACTACTTCTTGGCTTACAGCACCCCTTGGGGGTATTTTGACTTCAAC	2867
Oy	3081	AGATTTCAGTCCACTTTTCAACAGGTACGTGGGAGCACTCATCAACAAATTTGGGGGA	3140
Db	2868	AGATTTCAGTCCACTTTTCAACAGGTACGTGGCAAAAGACTCATCAACAACTGGGGGA	2927
Oy	3141	TTCCGGCCCAAGAGACTCAACTTCAACTCTTCAACATCCAGTCAAGAGAGTCAAGCAGAC	3200
Db	2928	TTCCGAACCAAGAGACTCAACTTCAAGCTCTTTTAAATTTCAAGTCAAAAGAGGTCAACGAG	2987
Oy	3201	AATGATGGCGTCAACAACATGCTTAATACCTTACAGACAGTTTCAAGTCTTCTCGGAC	3260

D	b	2988	AATGACGGTACGACGACGATTTGGCCAAATACCTTACACGACGGTTCAAGTGTATTACTGAC	3047
O	y	3261	TCGGAGTACCAAGCTTCCTCGTACGTCCTCGGCTCTGCGCACGAGGGCTGCTCCCTCGTTG	3320
D	b	3048	TCGGAGTACCAAGCTTCCTCGTACGTCCTCGGCTCTGCGGACATCAAGAGATGCTCCGCGCTTTC	3107
O	y	3321	CCGGCCGACCGTGTTCATGATTTCCGGCAATACGGCTACCTGACGGCTCAACAAATGACAGCTAA	3380
D	b	3108	CCACACAGACGCTTTCATAGTGTCCACAGATATGTAATCTCAACCTGAAACACGGGAGTCTAG	3167
O	y	3381	GCCTGGGAGCGTTCATCGCTCTTTTACTGCTGGCGAATATTTTCCCTTCACAGATGCTGAGAACG	3440
D	b	3168	GCAGTAGGACGCTCTTCACTTTTACTGCTGAGTACTTTCCTTCTCAAGATCTGGTATCC	3227
O	y	3441	GGCAACAACCTTTTACCTTCAGCTACACCTTTTGAGGAAGTCCCTTTTCAACGACGCTACGCG	3500
D	b	3228	GGAAACAACCTTTTACCTTCAGCTACACCTTTTGAGGACGTTCTCTTCCACAGCAGCTACGGCT	3287
O	y	3501	CACAGCCACAGACCTTGGACCCGGCTGATGATTCCTCTCATCGACCAATACCTGTATTACTTG	3560
D	b	3288	CACAGCCACAGATCTTGGACCCGCTCTGATGATTCCTCTCATCGACGACGTACCTGTATTACTTGG	3347
O	y	3561	AACAGAACTCAAAATCAAGTCCGGAAGTCCCAAAACAAGACGCTGTGTTTAAAGCGTGGG	3620
D	b	3348	AGCAAGAACAACTCTTCAAGTGAACCAACACGAGTCACAGCTTCAGTTTTTCTACGCC	3407
O	y	3621	TCCTCAGCTGATGTCTGTTCAGCCCAAAAACCTGGCTACCTGACCCCTGTATTCGGCAG	3680
D	b	3408	GGAGCGAGTGCATTTGGGACCAATCTTAAGAACTGGCTTCCTGACCTCTGTTAACGCCAG	3467
O	y	3681	CAGGCGCTTTTAAACAAACAAACACACACACACAACTTTTACTGGACGTGTGCT	3740
D	b	3468	CAGGAGATCTMAAGACATCTCGGATTAACAACAACAGTGAATCTGTGGACCTGAGCT	3527
O	y	3741	TCAAATATTAACCTCAATGGGCGGTGAATCCAATCAACCCCTGGCACCTGATGGCCCTCA	3800
D	b	3528	ACCAAGTACCACTTCAATAGGACAGACCTCTCTGGTGAATCCGGGCGCGGCATGACCAAC	3587
O	y	3801	CACAAAGACGACGAAACAAGTTCTTCCCATGACCGGTGTCAATGATTTTGTGAAAAAG	3860
D	b	3588	CACAAGACGATGAAGAAAAAGTTTTCCTCAGACCGGGGTTCATCTTTTGGGAAGCA	3647
O	y	3861	AGCGCCGAGCTTCAACACTGCTAGTGGACAAATGTCAATGATTACAGACGAAGAGAAATT	3920
D	b	3648	GGCTCAAGAAAAACAATGTGTGACATTTGAAAAAGTCAATGATTACAGACGAAGAGAAATC	3707
O	y	3921	AAAGCCACTAACCTCTGTGGCCACCGGAAAGATTTGGGACCGTGGACGTCAATTTCCAGAC	3980
D	b	3708	AGGACAAACCAATCCCGTGGCTACCGACACAGATAGTGTCTGTATTTACCAACTCCACAGGA	3767
O	y	3981	AGCAGCACAGACCTCTGGGACCGGAGATGTGCATCTATGGGAGCAATTAACCTGGCATGGTG	4040
D	b	3768	GGCAACAGACAGACGCTACCGGACGATGTCAACACACAAGGCGTTCTTCCAGGAGTGTCTC	3827
O	y	4041	TGGCAAGATAGACGCTGTACCTTGCAGAGGTCCATTTGGGCAAAATTTCTCACACAGAT	4100
D	b	3828	TGGCAGAGACAGAGATGTGTACTTCAAGGGGCCATCTGGGCAAGATTTCCACACAGGAC	3887
O	y	4101	GGACACTTTCAACCCCGTCTCTCTTTATGGGCGGCTTTTGAATCAGAAACCCGCTCTCTCAG	4160
D	b	3888	GGACATTTTCAACCCCTCTCCCTCATAGGTGGATTTGGACTTTAAACAACCTCTCTCCACAG	3947
O	y	4161	ATCTCATATCAAAAACAGCGCTGTCTCTGGCAATCTCTCGGCGGAGATTTTCAAGTACAAAG	4220
D	b	3948	ATTCTCATCAAGAACACCCCGGTACTCTGGCATCTTCTTGACCACTTCAGTGGCGCAAG	4007
O	y	4221	TTTTCCTTCAATTCATCACCCCAATATCTCCACAGGACAAATGTGAAATTTGAATGGGAG	4280
D	b	4008	TTTTCCTTCTTCAATCAACAGTACTTCCACGGGACAGGTCAAGCGTGGAGATGAGTGGGAG	4067
O	y	4281	CTGCAAGAAAGAAAACAGCAACCGCTGGAAATCCGGAATGCAATCAATCAATTAATGCA	4340
D	b	4068	CTGCAAGAAAGAAAACAGCAACCGCTGGAAATCCGGAATTAAGTACCTTCCCAACTCAAC	4127

LOCUS	AVU89790	4767 bp	DNA	circular	VRL 21-AUG-1997
DEFINITION	Adeno-associated virus 4, complete genome.				
VERSION	U89790.1				
KEYWORDS	GI:2337938				
SOURCE					
ORGANISM	adeno-associated virus 4.				
REFERENCE	adeno-associated virus 4. Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus. 1 (bases 1 to 4767)				
AUTHORS	Chlorini, J.A., Yang, L., Liu, Y., Safer, B. and Kotin, R.M.				
TITLE	Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles				
JOURNAL	J. Virol. 71 (9), 6823-6833 (1997)				
MEDLINE	97404695				
PUBMED	9261407				
REFERENCE	2 (bases 1 to 4767)				
AUTHORS	Chlorini, J.A., Yang, L., Kotin, R.M. and Safer, B.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-FEB-1997) MHR, NHLBI, NIH, Bld 10/7D18 10 Center Dr. MSC1654, Bethesda, MD 20982-1654, USA				
FEATURES	Location/Qualifiers				
source	1. 4767				
	/organism="adeno-associated virus 4"				
	/isolate="ATCC VR-646"				
	/db_xref="taxon:57579"				
	372. 2243				
	/codon_start=1				
	/product="rep"				
	/protein_id="AAC58044.1"				
	/db_xref="GI:2337939"				
	/translation="MPCGYEVLVVPSPDLDEHLPGISDSFYSVAWEKEMELPPSDDMDLNLBOAFLTAELKORFELVEMRVRSGAPALFFVPOREKDSYFHLILVETVGYKSMVAVRYSQIKELKLVTRIRGVEPOLPMPWAVTRTRNAGGKRVDDCITYNLPRKTPQBELVAMTMNDVYISACNLIAERKRLVAOHLTVSOTQONENONPNSDAVIVSSRTOBYAMELVGMLVDRIGTSEKQWIOEQDAYSIFNNAASNRSQIKALNDASIMELSTKTPADVLVGNQNPEDISNRIYRLLENGQVYDQAVSFMGAKQEGKNTVYLSAPATGKTNIIAELIAHNAVPFGCVAMNTNENPNCVCVRKMTVMEEGMTLAVSAAKALILGSKRVAVDQCKSSAIOIDPTPIVYTSNTMKCAVIDGNTSTPHOQPLDMRKPELITKRLAHDGKVTQKQVDFEFRRMASHVTEVHEFIVRKGARKRPAPDADISEPKACPSVAQPTSDAEAPVDYADRYQNKCSRHVMNMLPFCRCAGMNQNVDCIFHGWADCAACFPVSBESQPVSVVRKRTYQKLCPIHLHMGRAPEVACSAEELANVDDCDMEQ"				
	2260. 4464				
	/codon_start=1				
	/product="capsid"				
	/protein_id="AAC58045.1"				
	/db_xref="GI:2337940"				
	/translation="MTDGIQDLPWLEEDNLSEGVREWMALQGPAPRKANQOQHDNARGILVLPKYLGPNGNGLDGEPVNAADAALAEHDKAYDQOLKADNPLYLKNHADAFPOORLOGDSREGNIGRAVFOAKRVRLEPLIGVAGAGTAPEKRPILIESPOQPSNGIGRKGKQPAKKGLVPEDETGAQDPPRBSSTSGAMSDSEWRALAAAGAAVEGGQADVGNAASGDHMCDSVTSBEGHVTITSTRTWLPYTNHLVYKLGESLQSNYNGSTPWGFFDNRFCHGFSEBRDWORLINNMGKRPAMKPINIOGVKVTSSNGEYVANNLTSTVOLFADSYELPYWDAQOEGSLPEPNDVFMVPOGYGLVGTSSQOQDRAFLCYLEFFPSOMLRTGNNEFEITYSEKVPFHSVYAHOSLDRLAMPILDOYLMGLOSFTGTTLNAGTATNPETKLRPTNPSNFKRMNLPGPSIKOQSGKTNANQYKIPATSDSLIKETHSYTLDRGKMLTGPFPMAATGRPADSKRPSNQQLFAPGRKONGNTATYTGTLITFSEELAAATNADITDMWGNLPGGDQSNSNLPTVDRITLALGAVGAMQNRDLYEGGFWAKIPHTD				

BASE COUNT	1137 a	1405 c	1333 g	892 t	
ORIGIN	GHPSPSLIGFGLKHPDPIQNTLVNPNAPATPSTPNINSITQYSTQCVQVQID WEI0K5SKRMNBYQFTSTNSYNGQNSLMLNAPDAKXTERALGTRILVTLHNL"				
Query Match	55.1%;	Score 2598.6;	DB 14;	Length 4767;	
Best Local Similarity	74.6%;	Pred. No. 0;			
Matches 3584;	Conservative	0;	Mismatches 1094;	Indels 129;	Gaps 20;
QY	1	TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGTGGGAGCTTCGCGACCAAGAGTCCG	60		
DB	1	TTGGCCACTCCCTCTATAGCGCGCTCGCTCACTCACTCGGCTCTGGAGCCAAAGTCTCC	60		
QY	61	AGAGGCGAGAGCTTGTCTGTCCGCGCCCAACCGGCGAGCGAGGGCGGAGAGGAGTGT	120		
DB	61	AGACTGCGCGCTCTTGCGCGGCGAGGCGAGTGTAGCAGAGCGGATAGAGGAGTGT	120		
QY	121	GGCAACTCCATCACTAAGG-----GTAAATCGCGAAGCGCTCCAC	161		
DB	121	GCCAACTCCATCACTATAGTTTGCCCACTGACGTCATATGTACGCTTCAGGGTTAAGGAG	180		
QY	162	GCTGCGCGCTGAGCGCTGACGTAATTACGTATAGTGGAGTGTCTGTGA-----T	213		
DB	181	GTCCCTGTATTAGAGTACGCTAGTGTGCTATTTCCGCGAGCGTAGCGGCGCATACC	240		
QY	214	TAGCTGTACGTGA-----GTGCTTTGGGACATTTTGGGACCAAGTGGCCAT	263		
DB	241	AAGCTGCACGTCACAGCCACAGTGGTCCGTTTGGACAGTTTGGACCACTATGTGTAG	300		
QY	264	TTAGGGTATATATGGCGAGTGAAGCAGAGC-AGATCTCCATTTTGAACCGCAAAATTTGAA	322		
DB	301	G-AGGGATATATACCGGAGTGAAGCCAGCGAGAGTCTCATTTTGGCCGCAATTTTGAA	359		
QY	323	CGAGCAGCAGCCATCCGCGCTTCTACAGATCGTATCAAGTGGCCGAGCGACTTGAC	382		
DB	360	CGAGCAGCAGCATGCCGCGGTTCTACAGATCGTCTGAAGTGGCCAGCGACTTGAC	419		
QY	383	GAGCACTGCGCGGCAATTTCTGACTCGTTTGTAGCTGGGTGGCGGAGAGGAATGGAG	442		
DB	420	GAGCACTGCGCGGCAATTTCTGACTCTTTTGTAGCTGGGTGGCGGAGAGGAATGGAG	479		
QY	443	CTGCCCCCGGATTTGTGACATGATCTGTATCTGATTGAGCAGGACCCCTGACGCTGACC	502		
DB	480	CTGCCCCCGGATTTGTGACATGATCTGTATCTGATTGAGCAGGACCCCTGACGCTGACC	539		
QY	503	GAGAGCTGACGCGCATCTCTGTGTCATATGAGCGCGCGCTGAGTAAGGCCCGGAGGCC	562		
DB	540	GAAAGCTGACGCGCATCTCTGTGTCATATGAGCGCGCGCTGAGTAAGGCCCGGAGGCC	599		
QY	563	CTCTTCTTTTGTTCAGTTTGGAGAGGGGAGTCTTACTTCCACTTCCATATTTCTGTGGAG	622		
DB	600	CTCTTCTTTTGTTCAGTTTGGAGAGGGGAGTCTTACTTCCACTTCCATATTTCTGTGGAG	659		
QY	623	ACCAAGGGGGGCAATCCATGATGTCGTGGGCGCTTCCGATGTCAGATTAGGAGCAAGCTG	682		
DB	660	ACCGTGGCGTCATATCCATGTGTGTGGGCGCTTCTGATGTCAGATTAAAGAGAGCTGT	719		
QY	683	GTGACAGCACTTACCGCGGGATCGAGCGCAGCCCTGCGCAACTGTGTTGCGGATGACCAAG	742		
DB	720	GTGACCGGCACTTACCGCGGGATCGAGCGCAGCCCTGCGCAACTGTGTTGCGGATGACCAAG	779		
QY	743	ACGCGTATATGCGCGGAGGGGGAGCAAGTGTGTGACAGTGTCTACATCCCACTAC	802		
DB	780	ACGCGTATATGCGCGGAGGGGGAGCAAGTGTGTGACAGTGTCTACATCCCACTAC	839		
QY	803	CTCCGCGCAAGTCAAGCCGAGCTGAGTGAAGTGGCGTGAATTAACTAGGAGATATTA	862		
DB	840	CTGCTCCCAAGACCCAGCCGACTCTCAGTGGGCGTGAATTAATGACCAAGATATTA	899		
QY	863	AGCCCTGTTTGAACCTGCGAGCGGCAAAAGGCTGTGTGGCGAGCACTTGACCAAGCTC	922		
DB	900	AGCCCTGTTTGAATCTCGCGAGCTTAAACGCTGTGTGGCGAGCACTTACCGAAGCTG	959		

QY	923	AGCCGAGACCAGAGACGAGACAGAGAAATCTGAACCCCAATTCTTGACCGCGCTGTCAATC	982
Db	960	TCGCGAGACGCGAGGACGAGAACAGAGAAAAACAGAAACCCCAATTCTTGACCGCGCGGTCAATC	1019
QY	983	CGGTCAAAAAACCTCGCGCGGCTCAATGAGAGCTGTGGGTGGGTGGTGGAGACCGGGGCGATC	1042
Db	1020	AGGTCAAAAAACCTCGCGCGAGTCAATGAGAGCTGTGGGTGGGTGGTGGAGACCGCGGGATC	1079
QY	1043	ACCTCCGAGAAAGCACTGGATCCAGAGAGACAGAGGCTGTATCTATCTCTTGACAGCGCGCT	1102
Db	1080	ACGTCAGAAAAGCATGTGATCCAGAGAGGACAGGGGTCTCTATCTCTTGACAGCGCGCT	1139
QY	1103	TCCAACTGCGCGGTCCCAATGTAAGCGCGCTGTGACAATGCCGGCAAGATCATGGCGCTG	1162
Db	1140	TCCAACTGCGCGGTCACAAAATCAAGGCGCGCTGTGACAATGCTCTCCAAATCATGAGCTGT	1199
QY	1163	ACCAAATCCGGCGCGGACATCTGGTAAAGGCCCCGCTCGCGCCCGGAGACATTAATAACCAAC	1222
Db	1200	ACAAAGAAGGCTCCGAGCTACTGTGGTGGGACAGAACCGCGGAGAGACATTTCCAGCAAC	1259
QY	1223	CGCATCTAACCGCATCTGTGAGCTGAAACGAGCTTACGAACCTGACTACGCGGCTCCGCTTT	1282
Db	1260	CGCATCTAACCGAATCTCTGAGATGAAACGGGTACATCCGAGTACGCGGCTCCGCTTTC	1319
QY	1283	CTCGGCTGGGCGCCAGAAAAAGTTGGGGAAGCGCAACACATCTGGCTGTGTTGGCGCGCC	1342
Db	1320	CTGGGCTGGGCGCCAAAAAGAGTTGGGGAAGAGAACCATCTGGCTCTTTGGGCGCGCC	1379
QY	1343	ACCAAGGGCCAAACCAACATGGCGGAACCATCGGCCACGCCGTGCCCTTTCTAGCGGCTGC	1402
Db	1380	ACGAGGGGTAAAAACCAACATCGCGGAAGCATCGGCCACGCCGTGCCCTTTCTAGCGGCTGC	1439
QY	1403	GTCAACTGGAACCAATGAGAACTTTCCCTTCAATGATTTGCTGTGACAAAGATGGTATCTGG	1462
Db	1440	GTGAATGGAACCAATGAGAACTTTCCGTTCAACGATTTGCTGTGACAAAGATGGTATCTGG	1499
QY	1463	TGGAGAGAGGGCAAGATGACGGCCAAAGCTGTGAGTCCGCCAAGGCCATTTCTCGCGGCGC	1522
Db	1500	TGGAGAGAGGGCAAGATGACGGCCAAAGCTGTGAGAGAGCGCCAAAGGCCATCTTGGGCGGA	1559
QY	1523	AGCAAGGTGCGGTGTGGACCAAAAGTGCAGAGTGTCCGCCAATGTGACCCCAACCCCGTG	1582
Db	1560	AGCAAGGTGCGGTGTGGACCAAAAGTGCAGAGTATGGGCCAAGTGCACCAATCCCGTG	1619
QY	1583	ATCGTCACCTCCAACCAACCAACATGTGGCGCGGTGATTGACGGGAACAGCAACCACTTCCAG	1642
Db	1620	ATCGTCACCTCCAACCAACCAACATGTGGCGCGGTGATTGACGGGAATCTGACCACTTCCAG	1679
QY	1643	CACCAAGACCGGTTGCAGAGCCGGAATGTTCAAAATTTGAATCTCACCCGCGGTCTGGAGCAT	1702
Db	1680	CACCAACAAACCACTCCAGAGCCGGAATGTTCAAGTTGAGGTCTCACCAAGCGGCTGGAGCAC	1739
QY	1703	GACTTTGGCAGGTGACAAAGCAGGAAGTCAAAAGTTCTTCCGCTGGGCGCAGGATCAC	1762
Db	1740	GACTTTGGCAGGTGACCAAGCAGGAAGTCAAAAGCTTTTCCGCTGGGCGCTCAGATCAC	1799
QY	1763	GTGACCGAGGTGGCGCATGATGTTCTACGTCAAGAAAGGTGGAGGCCAAACAAAAGCCCGCC	1822
Db	1800	GTGACCGAGGTGATCAAGAGTTTATTCGTCAAGAAAGGTGGAGGCTGAAAAGAGGCGCGCC	1859
QY	1823	CCCGATGACGCGGGATMAAGCGAGGCCCAAGCGGGGCTCGCCCTCAGTGCAGGATCAATCG	1882
Db	1860	CCCATGACGCGAGATMAAGTAAAGTAAAGGCCCAAGCGGGGCTTTCGTCAGTTGGCCAGGCATCG	1919
QY	1883	ACGTCAAGCGGGAGAGAGCTCCGCTGGACTTTGCGCAAGGTACCAAAACAAATGTTCT	1942
Db	1920	ACGTCAAGCGCGGA--AGCTCCGCTGGACTTACGCGGACAGGTACCAAAACAAATGTTCT	1976
QY	1943	CGTACGGGGGAGATGCTCAGATGTGTTTCCCTCCAGACATGCGAGAGAAATGAATAG	2002
Db	1977	CGTACGGGGGATTAATCTGATGCTTTTCCCTCCGAGATGCGAGATGAATGATAG	2036
QY	2003	AATTTCACATTTGCTTACGACCGGAGCGAAGACTGTTCAGATGTGTTCCCGGCGCTG	2062

Db	2037	AATGGACATTGTGCTTACCGCAGGGGATCAGACCTGTGCCAGTGTCTCC---CGTG	2093
Qy	2063	TCAGAAATCTCAAC-----CGGTGCTGCAAGAAAGAGACGTAATCCGAATCTGTGCCATT	2116
Db	2094	TCGAAATCTCAACCCGTGTCTGTGTCTGAGAAAGGAGACGTATCAGAATCTGTGCCATT	2153
Qy	2117	CATCATCTGTGAGGGCGGGCTCCGAGATTGCTCTGAGCTGAGCTGTGATCAACGTG	2176
Db	2154	CATCATCATATGGGGAGGGCGCCGAGGTGACCTGTGCGCTGTGGACTGGCAATGTG	2213
Qy	2177	GACCTGGATGACTGTGTCTTCTGAGCAATAATGACTTAACCAAGTATGCTGCCGATG	2236
Db	2214	GACTTGGATGACTGTGACATGGAAACAATAATGACTCAACCAAGTATGACT--GACGG	2270
Qy	2237	TTATCTTCAGATTGGCTCGAGAGCAACCTCTCGAGGGGCAATTCGCGATGTGGGCACTT	2296
Db	2271	TTACTCTTCAGATTGGCTGAGAGCAACCTCTCGAAGGCGTTTCAAGTGTGTGGCGCT	2330
Qy	2297	GAAACTGTGAGCCCGGAAAGCCCAAGCCCAACAGCAAAAGCAGAGCAAGCGCGGGTCT	2356
Db	2331	GCAACTGTGAGCCCTTAACCCCAAGGCAAAATCAACAATCAGAGCAACGCTCGGGGTCT	2390
Qy	2357	GGTGCTTCTGTGCTACAAATACCTTGTGACCTTCAACGGAATCGACAAAGGGGAGCCGT	2416
Db	2391	TGTGCTTCTCGGGTTACAAATATCTTCGAGCCCGGCAACGAACTCGCAAGGGGGAAACCGT	2450
Qy	2417	CAAGCGCGGGAGACGAGAGCGGGCCCTCGAGCAAGCAAGGCCCTAACAGCAAGCAAGCTCAAGC	2476
Db	2451	CAAGCGACGGAGCGGGAGCCCTCGAGCAAGCAAGGCCCTAACAGCAAGCAAGCTCAAGC	2510
Qy	2477	GGGTGACAAATCCGTACTCTGCGGTATTAACACAGCGGACCGCGAGTTTCAGAGCGTCTGCA	2538
Db	2511	CGGTGACAAACCCCTAATCTCAAGTACAAACAAGCGGACCGGAGTTTCAGAGCGGCTTCA	2570
Qy	2537	AGAAATATCGTCTTTTGGGGGCAACCTTCGGGCGAGCACTTTCAAGGCAAAAGCGGT	2596
Db	2571	GGGCGACACATCGTTTGGGGGCAACCTTCGGGCGAGCACTTTCAAGGCAAAAGCGGT	2630
Qy	2597	TCTGCAACTCTCGGTCTGTGTTTGAGGAAGGGCGCTTAAGCGGCTCTCGAAAGAAACGTCC	2656
Db	2631	TCTTGAACCTCTTGTGTGTGTTTGAGGAAGGGCGGTGAAGCGGCTCTCGAAAGAAAGACCC	2690
Qy	2657	GGTAGAGCAGTGTGCGCAACAAGAGCCAGACTCTCTCTCGGCAATTCGCGCAAGACGCGCAGCA	2716
Db	2691	GTTGATTTGAATCCCGCCAGAGCCCGACTCTTCAAGCGGTATTCGCGCAAAAGGCAAGCA	2750
Qy	2717	GCCGCGTAAAGAGACTCAATTTTGTGACATGTGCGCACTCAGAGTCAATCCCGGATCC	2776
Db	2751	GCCGCGTAAAGAGACTCGTTT-----GGAAGACAAACTGGAGAGGCGGACGG	2801
Qy	2777	ACAACTCTCGAGAAACCTTCAGCAACCCCGCTGTGTGTGGAGCCTACTACAATGTCTTC	2836
Db	2802	ACCCCTTAGAGGATCACTTCCGAGGCATGTCTGATATACAG-----TGAGATGCCGTC	2855
Qy	2837	AGGCGGTGGCGCAACCAATGTGCGACAATAAAGAAAGCGCGCAGAGTGAATGCTTC	2896
Db	2866	AGCACTGTGGAGCTGCAAGTGTGAGGGCGGACAAAGTGTCCATGTGAGTGTGATGCTTC	2915
Qy	2897	AGAAATTTGCAATTGCGATTCCATGAGCTTGGGCGAGCAAGTCAATCAACACAGCACCCG	2956
Db	2916	GGGTGATTGGCATTTGCGATTCCACTGTGTGAGGGCCACGTCAACACACAGCACAG	2975
Qy	2957	CACCTGTGGGCTTTGCCCACTTCAATTAACACTCTTCAAGCAAAATTTCTCAAGTCTTCAAC	3016
Db	2976	AACCTGGGTCTTGGCCCACTTCAACAACAACAACCTTCAAGCAAGCACTCGGA-----	3024
Qy	3017	GGGGGCGCAGCAACGACCACTTCTTGGCTAAGACACCCCTGTGGGGGTATTTGATTT	3076
Db	3025	-GAGAGCTGTGACATCAACCTTCAACGGAATTTTCAACCCCTGTGGGGATTTGACTT	3083
Qy	3077	CAACGATTCACATGCACTTTTCAACGATGACTGTGCGAGCACTCAACAACAATG	3136

Db 3084 CAACCGCTCCACTGCACTTCTCACCAGTACGAGGAGCTCATCAACAACACTG 3143
Qy 3137 GGGATTCGGGCGGAGAGACTCAACTTCAAACTTCAAACTTCAAACTTCAAACTTCA 3196
Db 3144 GGGATTCGGGCGGAGAGACTCAACTTCAAACTTCAAACTTCAAACTTCAAACTTCA 3203
Qy 3197 GACGAATGATGAGCTTCAACAACATGCTAATTAACCTTACAGAGAGGTTCAAGCTTCTC 3256
Db 3204 GACCTCGAAGCGGAG 3263
Qy 3257 GAGCTCGAGAGTACAGAGCTTCCGTAAGTCTCGAGTCTGAGCAAGAGAGAGAGAGAG 3316
Db 3264 GAGCTCGAGAGTACAGAGCTTCCGTAAGTCTCGAGTCTGAGCAAGAGAGAGAGAGAG 3323
Qy 3317 GTTCCCGGCGGAGAGTCTTCAATGATTCGCAATAGGCTAC-----CTGAGCTCAA 3367
Db 3324 TTTTCCCAACGAGCTTCTTAAAGTGGCCGAGTACGCTACGCTGAGAGTGGAGCGGAA 3383
Qy 3368 CAATGAGAGCGCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3427
Db 3384 CACTTCGAG 3443
Qy 3428 GATGCTGAG 3487
Db 3444 GATGCTGAG 3503
Qy 3488 CAG 3547
Db 3504 CTGAGATGAG 3563
Qy 3548 CTTGATTAATCTGAG 3604
Db 3564 CTTGATTAATCTGAG 3623
Qy 3605 GCTGTTTACCGTGGGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3664
Db 3624 CAATCTTAATCAAG 3683
Qy 3665 ACCGTGTTATGAG 3724
Db 3684 GCTTCAATCAAG 3743
Qy 3725 TACCTGAG 3769
Db 3744 CACGGGAG 3803
Qy 3770 CATCATCAAGCTGAG 3829
Db 3804 TGCCCTGAG 3860
Qy 3830 CATGAG 3889
Db 3861 CAAG 3920
Qy 3890 CAATGTCATGATTAAG 3949
Db 3921 GACTCTGATCTTCACTCTGAG 3980
Qy 3950 ATTGAG 4009
Db 3981 GTGGGAG 4040
Qy 4010 GCATGCTATGAG 4069
Db 4041 GACAG 4100
Qy 4070 TCCGATTTGAG 4129
Db 4101 TCCGATTTGAG 4160
Qy 4130 CGGCTTTGAG 4189
Db 4161 TGGGTTTGGGCTGAG 4220

Qy 4190 GAATCCCGGCGGAG 4249
Db 4221 GAATCCCGGCGGAG 4280
Qy 4250 AGGAG 4309
Db 4281 TGGGAG 4340
Qy 4310 TCCGAG 4369
Db 4341 CCGGAG 4400
Qy 4370 CAACAAATGAG 4429
Db 4401 TGGGAG 4460
Qy 4430 GTAATTAAG 4488
Db 4461 GTAATTAAG 4520
Qy 4489 TCCCTCTTATCTTATC-GGTTACATGAGTATAGCTTACATTAAGTCTG-----CTTGG 4541
Db 4521 TCCCTCTTATCTTATC-GGTTACATGAGTATAGCTTACATTAAGTCTG-----CTTGG 4580
Qy 4542 TGGGCTTGGGAG 4591
Db 4581 TGGGCTTGGGAG 4640
Qy 4592 AGTTGCCACTTCCCTCTTCTGCGGCTGCGTCTGCGTCTGCGTCTGCGTCTGCGTCTGCGTCT 4651
Db 4641 AGTTGCCACTTCCCTCTTCTGCGGCTGCGTCTGCGTCTGCGTCTGCGTCTGCGTCTGCGTCT 4700
Qy 4652 GCAG 4711
Db 4701 CAG 4760
Qy 4712 TGGGCAA 4718
Db 4761 TGGGCAA 4767

Search completed: July 20, 2003, 13:19:53
Job time : 12464.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:53:46 ; Search time 32.1207 Seconds*
(without alignments)
10025.800 Million cell updates/sec

Title: US-09-807-802a-1_COPY_4576_4718

Perfect score: 143

Sequence: 1 ttaccctagtgatgagtc.....cgcagagagggagtgggcaaa 143

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Genseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	143	100.0	4718	21	AAD00772
2	125	87.4	143	21	Adeno-associated v
3	121.8	85.2	4683	22	AAV6 DNA sequence
4	119	83.2	143	21	AAD00781
5	119	83.2	4718	21	Adeno-associated v
6	114.8	80.3	4683	22	AAV6 DNA sequence
7	107.4	75.1	145	18	Inverted terminal
8	107.4	75.1	145	24	Adeno-associated v
9	107.4	75.1	165	15	AAO66769
					Double-D ITR termi

10	107.4	75.1	165	18	AAV94462	Adeno associated v
11	107.4	75.1	165	22	AAD03535	Inverted terminal
12	107.4	75.1	192	16	AAV03384	Right-hand invert
13	107.4	75.1	272	21	AAV47166	Head-to-tail ITR s
14	107.4	75.1	272	22	AAD03534	AAV circular inter
15	107.4	75.1	4675	22	AAH41481	Adeno-associated v
16	107.4	75.1	4675	22	AAV89931	Nucleotide sequenc
17	107.4	75.1	4675	22	ABA02989	Adeno-associated v
18	107.4	75.1	4679	22	AAV66972	Adeno-associated v
19	107.4	75.1	4679	22	AAV23750	AAV2 DNA sequence
20	107.4	75.1	4680	17	AAV09008	Wild-type adeno-as
21	107.4	75.1	8698	22	AAH26325	Wild-type adeno-as
22	105.8	74.0	189	24	AAV17057	AAV terminal repa
23	105.8	74.0	191	24	AAV58791	AAV 5' ITR from AAV
24	105.8	74.0	232	21	AAV47164	Head-to-tail ITR s
25	105.8	74.0	272	22	AAD03532	AAV circular inter
26	105.4	73.7	174	21	AAV10404	Human adeno-associ
27	102.6	71.7	272	21	AAV47166	Head-to-tail ITR s
28	102.6	71.7	272	22	AAD03534	AAV circular inter
29	101	70.6	174	24	AAV58793	AAV 3' ITR from AAV
30	101	70.6	183	21	AAV10405	Human adeno-associ
31	99.4	69.5	145	14	AAV41448	AAV2 inverted term
32	99.4	69.5	145	16	AAV03385	Strict inverted te
33	99.4	69.5	145	20	AAV34295	Adeno-associated v
34	99.4	69.5	165	15	AAO66769	Double-D ITR termi
35	99.4	69.5	165	18	AAV49462	Adeno associated v
36	99.4	69.5	165	22	AAD03535	Inverted terminal
37	99.4	69.5	955	24	AAD37254	Adeno-associated v
38	99.4	69.5	955	24	AAD37254	Adeno-associated v
39	99.4	69.5	987	24	AAD37261	Adeno-associated v
40	99.4	69.5	987	24	AAD37261	Adeno-associated v
41	99.4	69.5	4414	24	AAD37260	Adeno-associated v
42	99.4	69.5	4476	24	AAD37259	Adeno-associated v
43	99.4	69.5	4476	24	AAD37259	Adeno-associated v
44	99.4	69.5	4476	24	AAD37259	Adeno-associated v
45	99.4	69.5	4498	24	AAD37258	Adeno-associated v

ALIGNMENTS

RESULT 1	
AAD00772	standard; DNA; 4718 BP.
ID	AAD00772
XX	08-SEP-2000 (first entry)
AC	Adeno-associated virus serotype 1 DNA.
XX	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein; cap protein; recombinant viral vector; gene delivery; gene therapy; vaccine; transgene; ss.
XX	Adeno-associated virus serotype 1.
OS	Adeno-associated virus serotype 1.
XX	Key
XX	repeat_unit
XX	Location/Qualifiers
XX	1..143
XX	/tag= a
XX	/label= 5' ITR
XX	/note= "Inverted terminal repeat which is capable of forming T-shaped hairpin structure"
XX	89..110
XX	/tag= b
XX	/bound moiety= "Rep protein"
XX	124..125
XX	/tag= c
XX	/note= "Terminal resolute site (TRS)"
XX	219..226
XX	/tag= d
XX	/bound moiety= "USP"
XX	/note= "B box"

FT	promoter..299	/+tag= e
FT		/label= p5_promoter
FT	protein_bind	237..245
FT		/+tag= f
FT		/bound_moiety= "YYI factor"
FT	TATA_signal	270..275
FT		/+tag= g
FT		/label= p5_TATA-Box
FT	misc_feature	299..306
FT		/+tag= h
FT		/note= "Y11/p5 RNA"
FT	CDS	335..2206
FT		/+tag= i
FT		/function= "Reg 78"
FT		/product= "regulates replication and integration
FT		of AAV DNA into host cell's chromosome"
FT	CDS	335..2272
FT		/+tag= j
FT		/product= "Rep 68"
FT		/function= "regulates replication and integration
FT		of AAV DNA into host cell's chromosome"
FT		/note= "The coding region is interrupted by intron
FT	intron	1924..2220
FT		/+tag= k
FT		/note= "This region interrupts the coding sequence
FT		of Rep 68 and Rep 40"
FT	TATA_signal	857..862
FT		/+tag= l
FT		/label= p19_TATA_Box
FT	misc_feature	882..883
FT		/+tag= m
FT		/note= "P19 RNA"
FT	CDS	1007..2206
FT		/+tag= n
FT		/product= "Rep 52"
FT		/function= "regulates replication and integration
FT		of AAV DNA into host cell's chromosome"
FT	CDS	1007..2272
FT		/+tag= o
FT		/product= "Rep 40"
FT		/function= "regulates replication and integration
FT		of AAV DNA into host cell's chromosome"
FT		/note= "The coding region is interrupted by intron
FT	TATA_signal	1836..1841
FT		/+tag= p
FT		/label= P40_TATA-BOX
FT	misc_feature	1875..1876
FT		/+tag= q
FT		/note= "P40 RNA"
FT	CDS	2223..4433
FT		/+tag= r
FT		/product= "VP1 protein"
FT		/note= "Capsid protein"
FT	CDS	2634..4433
FT		/+tag= s
FT		/product= "VP2 protein"
FT		/note= "Capsid protein"
FT		/partial
FT	CDS	2829..4433
FT		/+tag= t
FT		/product= "VP3 protein"
FT		/note= "Capsid protein"
FT	polyA_signal	4447..4452
FT		/+tag= u
FT	repeat_unit	4576..4718
FT		/+tag= v
FT		/label= 3..ITR
FT		/note= "Inverted terminal repeat which is capable
XX		of forming T-shaped hairpin-structure"
XX	MO200028061-A2.	

PD	18-MAY-2000.	
XX		
PF	02-NOV-1999;	99WO-US25694.
XX		
FR	05-NOV-1998;	98US-0107114.
XX		
PA	(UYPE-) UNIV PENNSYLVANIA.	
XX		
PI	Wilson JM, Xiao W;	
XX		
DR	WPI: 2000-376571/32	
XX		
DR	P-ESDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,	
XX		
XX	AAY71169.	
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for	
PT	preparation of medicament for delivery of a transgene to a host	
XX		
XX	Claim 1; Fig 1; 108pp; English.	
XX		
SS		
CC	The present sequence is an adeno-associated virus serotype 1 (AAV-1)	
CC	DNA characterised by two inverted terminal repeats (ITR) and open	
CC	reading frames for rep and capsid (cap) proteins. The rep reading frame	
CC	encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap	
CC	reading frame encodes three structural proteins, VP1, VP2 and VP3.	
CC	The AAV-1 sequence or its fragments particularly ITRs, rep and cap	
CC	coding regions, are useful in production of recombinant viral vectors	
CC	for gene delivery. These vectors can be used as gene therapy	
CC	vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does	
CC	not induce the formation of neutralising antibodies specific to any	
CC	serotype of AAV hence is useful for transforming host cells, and in the	
CC	preparation of a medicament for the delivery of transgene to a host.	
XX		
XX		
SO	Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;	
	Query Match	100.0%; Score 143; DB 21; Length 4718;
	Best Local Similarity	100.0%; Pred. No. 1,8e-28;
	Matches 143; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 TTACCCCTGAGTGAAGAGCTTCCCATCTCTCTGCGCGCTCGCTCGCTGGTGGGCC	60
DB	4576 TTACCCCTGAGTGAAGAGCTTCCCATCTCTCTGCGCGCTCGCTCGCTGGTGGGCC	4635
QY	61 TGGGACCAAAAGTCCGACAGCGAGAGCTCTGCTGCGGCGCCACGAGCGAGCGA	120
DB	4636 TGGGACCAAAAGTCCGACAGCGAGAGCTCTGCTGCGGCGCCACGAGCGAGCGA	4635
QY	121 GCGCGCAGAGAGGGAGTGGCGAA	143
DB	4696 GCGCGCAGAGAGGGAGTGGCGAA	4718
	RESULT 2	
	AAD00781	
ID	AAD00781 standard; DNA; 143 BP.	
XX		
AC	AAD00781;	
XX		
DT	08-SEP-2000 (first entry)	
XX		
XX		
DE	Adeno-associated virus serotype 1 5' inverted terminal repeat.	
XX		
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;	
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;	
KW	vaccine; transgene; inverted terminal repeat; ITR; ss.	
XX		
OS	Adeno associated virus serotype 1.	
XX		
PH	Key	Location/Qualifiers
FT	misc_binding	1..41
FT		/*tag= a
FT		/bound_moiety= "binds to nucleotides 85..125"
FT	stem_loop	42..62
FT		/*tag= b

```
FT      stem_loop      64..84
FT      /*cag= c
FT      misc_binding    85..125
FT      /*tag= d
FT      /bound_molety= "binds to nucleotides 41..1"
XX
XX      WO200028061-A2.
XX
XX      18-MAY-2000.
XX
XX      02-NOV-1999;    99WO-US25694.
XX
XX      05-NOV-1998;    98US-0107114.
XX
XX      (UNIV ) UNIV PENNSYLVANIA.
XX
XX      Wilson JM, Xiao W;
XX
XX      WPI; 2000-376571/32.
XX
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host
XX
XX      Example 2; Fig 2; 108pp; English.
XX
XX      The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX      which is characterised by two inverted terminal repeats (ITR) and open
XX      reading frames for rep and capsid (cap) proteins. The rep reading frame
XX      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX      reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX      coding regions, are useful in production of recombinant viral vectors
XX      for gene delivery. These vectors can be used as gene therapy
XX      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX      not induce the formation of neutralising antibodies specific to any
XX      serotype of AAV hence is useful for transforming host cells, and in the
XX      preparation of a medicament for the delivery of transgene to a host.
XX      The present sequence is a 5' ITR of AAV-1 DNA which is useful in the
XX      production of recombinant viral vector. The ITR forms a T-shaped hairpin
XX      structure.
XX
XX      Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;
XX
XX      Query Match      87.4%; Score 125; DB 21; Length 143;
XX      Best Local Similarity 100.0%; Pred. No. 7.6e-24;
XX      Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      19 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 78
XX      DB      1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 60
XX
XX      QY      79 AGACGGCAGAGCTGTGCTTGGCGGCCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 138
XX      DB      61 AGACGGCAGAGCTGTGCTTGGCGGCCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 120
XX
XX      QY      139 GGCAA 143
XX      DB      121 GGCAA 125
XX
XX
XX
XX      RESULT 3
XX      AAF23749/c
XX      ID AAF23749 standard; DNA; 4683 BP.
XX
XX      AAF23749;
XX
XX      28-MAR-2001 (first entry)
XX
XX      AAV6 DNA sequence.
XX
XX      AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX      atherosclerosis; sickle cell anaemia; thalassemia;
XX      blood clotting disorder; diabetes; ss.
```

```
XX      Adeno associated virus.
XX
XX      US6156303-A.
XX
XX      05-DEC-2000.
XX
XX      11-JUN-1997;    97US-0873168.
XX
XX      11-JUN-1997;    97US-0873168.
XX
XX      (UNIV ) UNIV WASHINGTON.
XX
XX      Russell DM, Rutledge EA;
XX
XX      WPI; 2001-060164/07.
XX
XX      Adeno-associated virus serotype 6 and viral vector derived from it for
XX      gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX      syndrome, sickle cell anemia, thalassemia and diabetes
XX
XX      Claim 1; Fig 1; 50pp; English.
XX
XX      The present invention relates to adeno-associated virus serotypes. The
XX      present sequence is the DNA sequence of one such serotype (AAV6). AAV6
XX      can be used to construct AAV viral vectors for use in gene therapy for a
XX      range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX      sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
XX      The AAV viral vectors have increased transduction efficiency of a
XX      particular host cell as the AAV virion containing the AAV vector genome
XX      can be modified to express a capsid protein of an AAV serotype that
XX      transduces the selected host cell.
XX
XX      Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
XX
XX      Query Match      85.2%; Score 121.8; DB 22; Length 4683;
XX      Best Local Similarity 98.4%; Pred. No. 6.1e-23;
XX      Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX      QY      19 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 78
XX      DB      4683 TTGGCCACTCCCTCTAAGCCCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 4624
XX
XX      QY      79 AGACGGCAGAGCTGTGCTTGGCGGCCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 136
XX      DB      4623 AGACGGCAGAGCTGTGCTTGGCGGCCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 4564
XX
XX      QY      139 GGCAA 143
XX      DB      4563 GGCAA 4559
XX
XX
XX
XX      RESULT 4
XX      AAD00781/c
XX      ID AAD00781 standard; DNA; 143 BP.
XX
XX      AAD00781;
XX
XX      08-SEP-2000 (first entry)
XX
XX      Adeno-associated virus serotype 1 5' inverted terminal repeat.
XX
XX      Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX      cap protein; recombinant viral vector; gene delivery; gene therapy;
XX      vaccine; transgene; inverted terminal repeat; ITR; ss.
XX
XX      Adeno associated virus serotype 1.
XX
XX      Key      Location/Qualifiers
XX      FT      misc_binding    1..41
XX      FT      /*cag= a
XX      FT      /bound_molety= "binds to nucleotides 85..125"
XX
XX      stem_loop    42..62
```

```
FT      stem_loop      /*tag= b
XX      64..84
FT      /*tag= c
FT      misc_binding    85..125
FT      /*tag= d
FT      /bound_molecety= "binds to nucleotides 41..1"
XX
XX      WO200028061-A2.
XX      18-MAY-2000.
XX
XX      02-NOV-1999; 99WO-US25694.
XX
XX      05-NOV-1998; 98US-0107114.
XX
XX      (TYPE-) UNIV PENNSYLVANIA.
XX
XX      Wilson JM, Xiao W;
XX
XX      WPI: 2000-376571/32.
XX
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host
XX
XX      Example 2; Fig 2; 108pp; English.
XX
XX      The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX      which is characterised by two inverted terminal repeats (ITR) and open
XX      reading frames for rep and capsid (cap) proteins. The rep reading frame
XX      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX      reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX      coding regions, are useful in production of recombinant viral vectors
XX      for gene delivery. These vectors can be used as gene therapy
XX      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX      not induce the formation of neutralising antibodies specific to any
XX      serotype of AAV hence is useful for transforming host cells, and in the
XX      preparation of a medicament for the delivery of transgene to a host.
XX      The present sequence is a 5' ITR of AAV-1 DNA which is useful in the
XX      production of recombinant viral vector. The ITR forms a T-shaped hairpin
XX      structure.
XX
XX      Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;
XX
XX      Query March 83.2%; Score 119; DB 21; Length 143;
XX      Best Local Similarity 89.5%; Pred. No. 2.8e-22;
XX      Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
XX
XX      QY      1 TTACCCCTAGTATGAGTGTCCCACTCCCTCTGTGCGGCTCGCTCGTGGAGGCC 60
XX      Db      143 TTACCCCTAGTATGAGTGTCCCACTCCCTCTGTGCGGCTCGCTCGTGGAGGCC 84
XX
XX      QY      61 TGGGACCAAGAGTCCGCAACGGCAGAGCTCTGTGCGGCCCAACCGAGGAGCGA 120
XX      Db      83 GGCAGAGAGAGAGCTCTGCGGACCTTTGGTCCGCAAGGCCCAACCGAGGAGCGA 24
XX
XX      QY      121 GCGGCGAGAGAGGAGTGGGCAA 143
XX      Db      23 GCGGCGAGAGAGGAGTGGGCAA 1
XX
XX      RESULT 5
XX      AAD00772/c
XX      ID      AAD00772 standard; DNA; 4718 BP.
XX
XX      AC      AAD00772;
XX
XX      DT      08-SEP-2000 (first entry)
XX
XX      DE      Adeno-associated virus serotype 1 DNA.
XX
XX      XX      Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX      KW      cap protein; recombinant viral vector; gene delivery; gene therapy;
```

```
KW      vaccine; transgene; ss.
XX
XX      Adeno associated virus serotype 1.
XX      OS
XX      Key
XX      repeat_unit 1..143
XX      /tag= a
XX      /label= 5' ITR
XX      /note= "Inverted terminal repeat which is capable
XX      of forming T-shaped hairpin structure"
XX      89..110
XX      /tag= b
XX      /bound_molecety= "Rep protein"
XX      124..125
XX      /tag= c
XX      /note= "terminal resolute site (TRS)"
XX      219..226
XX      /tag= d
XX      /bound_molecety= "USF"
XX      236..239
XX      /tag= e
XX      /label= P5_promoter
XX      237..245
XX      /tag= f
XX      /bound_molecety= "Y1 factor"
XX      270..275
XX      /tag= g
XX      /label= P5_TATA_Box
XX      299..306
XX      /tag= h
XX      /note= "Y1/P5 RNA"
XX      335..2206
XX      /tag= i
XX      /product= "Rep 78"
XX      /function= "regulates replication and integration
XX      of AAV DNA into host cell's chromosome"
XX      335..2272
XX      /tag= j
XX      /product= "Rep 68"
XX      /function= "regulates replication and integration
XX      of AAV DNA into host cell's chromosome"
XX      1924..2220
XX      /tag= k
XX      /note= "This region interrupts the coding sequence
XX      of Rep 68 and Rep 40"
XX      857..862
XX      /tag= l
XX      /label= P19_TATA_Box
XX      882..883
XX      /tag= m
XX      /note= "P19 RNA"
XX      1007..2206
XX      /tag= n
XX      /product= "Rep 52"
XX      /function= "regulates replication and integration
XX      of AAV DNA into host cell's chromosome"
XX      1007..2272
XX      /tag= o
XX      /product= "Rep 40"
XX      /function= "regulates replication and integration
XX      of AAV DNA into host cell's chromosome"
XX      1836..1841
XX      /tag= p
XX      /label= P40_TATA_BOX
XX      1875..1876
XX      /tag= q
XX      /note= "P40 RNA"
XX      2223..4433
XX      /tag= r
XX      /product= "VP1 protein"
XX
XX      TATA_signal
XX
XX      misc_feature
XX
XX      CDS
XX
XX      CDS
```

```
FT FT /note= "Capsid protein"
FT FT 2634..4433
FT FT /*tag= s
FT FT /product= "VP2 protein"
FT FT /note= "Capsid protein"
FT FT /partial
FT CDS 2829..4433
FT FT /*tag= t
FT FT /product= "VP3 protein"
FT FT /note= "Capsid protein"
FT FT 4447..4452
FT FT /*tag= u
FT FT 4576..4718
FT FT /*tag= v
FT FT /label= 3' ITR
FT FT /note= "Inverted terminal repeat which is capable
FT FT of forming T-shaped hairpin structure"
FT
XX MO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX PF 02-NOV-1999; 99WO-US25694.
XX
XX PR 05-NOV-1998; 98US-0107114.
XX
XX (UNPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Xiao W;
XX
XX WPI; 2000-376571/32.
XX
XX P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
XX AAY71169.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
XX
XX Claim 1; Fig 1; 108bp; English.
XX
XX The present sequence is an adeno-associated virus serotype 1 (AAV-1)
XX characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX
XX Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
SQ
XX
XX Query Match 83.2%; Score 119; DB 21; Length 4718;
XX Best Local Similarity 89.5%; Pred. No. 3.3e-22;
XX Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
XX
XX 1 TTATCCCTAGTATGATGAGTTGCCACTCCTCTGTGCGCGCTGCTGCTGCGGGCC 60
XX 143 TTATCCCTAGTATGAGTTGCCACTCCTCTGTGCGCGCTGCTGCTGCGGGCC 84
XX
XX 61 TGGGGAACCAAGTCCGACAGCGGAGAGCTGCTGCGCGCGCCACCGAGGAGCGA 120
XX 83 GGCAGACAGAGCTCTGCGCGCTGCGGACCTTTGTCTCCGAGGCCACCGAGGAGCGA 24
XX
XX 121 GCGGCGAGAGGAGGAGTGGCAA 143
XX 23 GCGGCGAGAGGAGGAGTGGCAA 1
XX
XX RESULT 6
XX AAF23749
```

```
ID AAF23749 standard; DNA; 4683 BP.
XX
XX AAF23749;
AC
XX 28-MAR-2001 (first entry)
DT
XX
XX AAV6 DNA sequence.
DE
XX
XX AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX atherosclerosis; sickle cell anaemia; thalassaemia;
XX blood clotting disorder; diabetes; ss.
XX
XX Adeno associated virus.
OS
XX
XX US6156303-A.
PN
XX
XX 05-DEC-2000.
PD
XX
XX 11-JUN-1997; 97US-0873168.
PF
XX
XX 11-JUN-1997; 97US-0873168.
PR
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Russell DW, Rutledge EA;
XX
XX WPI; 2001-060164/07.
XX
XX Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX
XX Claim 1; Fig 1; 50bp; English.
XX
XX The present invention relates to adeno-associated virus serotypes. The
XX present sequence is the DNA sequence of one such serotype (AAV6). AAV6
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX
XX Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
SQ
XX
XX Query Match 80.3%; Score 114.8; DB 22; Length 4683;
XX Best Local Similarity 88.0%; Pred. No. 4.1e-21;
XX Matches 125; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
XX
XX 2 TACCCCTAGTATGAGTTGCCACTCCTCTGTGCGCGCTGCTGCTGCGGGCCCT 61
XX 4542 TACCCCTAGTATGAGTTGCCACTCCTCTATGCGGCTGCTGCTGCGGGCCG 4601
XX
XX 62 GCGGACCAAGTCCGACAGCGGAGAGCTGCTGCTGCGCGCCACCGAGGAGCGAG 121
XX 4602 GCGAGACAGAGCTCTGCGCGCTGCGGACCTTTGTCTCCGAGGCCACCGAGGAGCGAG 4661
XX
XX 122 CCGGCGAGAGGAGGAGTGGCAA 143
XX 4662 CCGGCATAGAGGAGGAGTGGCAA 4683
XX
XX RESULT 7
XX AAT63408/c
ID AAT63408 standard; DNA; 145 BP.
XX
XX AAT63408;
AC
XX
XX 18-DEC-1997 (first entry)
DT
XX
XX Inverted terminal repeat from adeno associated virus serotype 2.
XX
```


KW Inverted terminal repeat; ITR; wild type; adeno associated virus;
 KW AAV; serotype 2; Rep binding site; integration system; treatment;
 KW acquired; inherited; disease; gene therapy; ss.
 XX
 OS Adeno associated virus.
 XX
 FH Key Location/Qualifiers
 FT repeat_unit. 1..42
 FT /tag= a
 FT /rpt_type= inverted
 FT /note= "region A"
 FT repeat_unit 85..125
 FT /tag= b
 FT /rpt_type= inverted
 FT /note= "region A"
 FT repeat_unit 43..51
 FT /tag= c
 FT /rpt_type= inverted
 FT /note= "region C"
 FT repeat_unit 53..61
 FT /tag= d
 FT /rpt_type= inverted
 FT /note= "region C"
 FT repeat_unit 64..73
 FT /tag= e
 FT /rpt_type= inverted
 FT /note= "region B"
 FT repeat_unit 75..84
 FT /tag= f
 FT /rpt_type= inverted
 FT /note= "region B"
 FT misc_binding 20..31
 FT /tag= g
 FT /label= Rep_binding_site
 FT /note= "region B"
 FT misc_feature 126..145
 FT /tag= h
 FT /note= "region D"
 FT
 FT
 PN WO9709442-A1.
 XX
 PD 13-MAR-1997.
 XX
 XX
 XX 06-SEP-1996; 96WO-US14312.
 XX
 XX 08-SEP-1995; 95US-0525866.
 XX
 PA (AVIG-) AVIGEN INC.
 XX
 PI Natsoulis G, Surosky RT;
 XX
 DR WPI; 1997-192917/17.
 XX
 FT Nucleic acid construct for integration into target site in recipient
 FT genome - comprises target sequence to bind AAV Rep and heterologous
 FT nucleotide sequence, for treatment of acquired or inherited disease
 XX
 PS Claim 3; Fig 1; 54pp; English.
 XX
 CC The present sequence is the inverted terminal repeat (ITR) from the
 CC wild type adeno associated virus (AAV) serotype 2 genome.
 CC A novel nucleic acid construct, comprises a targeting sequence
 CC capable of binding AAV Rep, e.g. the present sequence, and at least
 CC 1 heterologous nucleotide sequence arranged relative to the
 CC targeting sequence, so that it is integrated into a target site in
 CC a recipient genome. The integration system can be used to integrate
 CC a selected nucleotide sequence into a recipient cell genome, useful
 CC in the treatment of an acquired or inherited disease, e.g. long
 CC term gene therapy. The integration system has the site specific
 CC integration characteristics of AAV, is not limiting in regard to
 CC the size of the nucleotide sequence to be delivered and does not
 CC result in the integration of viral gene sequences into the
 CC recipient cell genome. Specifically it can be used to safely and
 CC efficiently deliver, and integrate large nucleotide sequences into

CC a suitable recipient cell genome without the risk of causing
 CC insertional mutagenesis due to random integration events.
 XX
 XX Sequence 145 BP; 20 A; 52 C; 49 G; 24 T; 0 other;
 SO
 Query Match 75.1%; Score 107.4; DB 18; Length 145;
 Best Local Similarity 85.1%; Pred. No. 3e-19;
 Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
 QY 3 ACCCTGATGATGAGTGGCCACTCCCTCTCGCGGCTCGCTCGGTGGGCGCTG 62
 DB 141 ACCCTGATGATGAGTGGCCACTCCCTCTCGCGGCTCGCTCGGTGAGGCGG 82
 QY 63 CGACCAAGATCGCGAGAGCGAGAGCTGCTGTCGCCGCCACCGAGCGAGC 122
 DB 81 GCGACCAAGATCGCGAGAGCGAGAGCTGCTGTCGCCGCCCTGATGAGCGAGC 22
 QY 123 GCGCAGAGAGGAGTGGGCA 143
 DB 21 GCGCAGAGAGGAGTGGGCA 1
 RESULT 8
 ABA02990 standard; DNA; 145 BP.
 XX
 XX ABA02990;
 XX
 XX 19-FEB-2002 (first entry)
 DT
 XX
 DE Adeno-associated virus 2 ITR SEQ ID NO 2.
 KW
 KW Cytostatic; virucide; apoptosis inducer; p53; Saos-2 cell; infection;
 KW cancer; virus; HPV16; HPV18; adeno-associated virus 2; AAV-2; ss.
 XX
 OS Adeno associated virus 2.
 XX
 FH Key Location/Qualifiers
 FT misc_structure 1..145
 FT /tag= a
 FT /label= "ITR"
 FT misc_feature 72
 FT /tag= b
 FT /note= "unpaired base"
 FT misc_feature 94
 FT /tag= c
 FT /note= "unpaired base"
 FT
 FT
 PN WO200180840-A2.
 XX
 PN 01-NOV-2001.
 XX
 XX 20-APR-2001; 2001WO-GB01795.
 XX
 XX 20-APR-2000; 2000GB-0009887.
 XX
 PA (BTGI-) BTG INT LTD.
 XX
 PI Raj K, Beard PM;
 XX
 DR WPI; 2002-041365/05.
 XX
 FT Single stranded and/or looped DNA for treating mutant p53 associated
 FT cancer or infection that inhibit cellular p53, having portion with an
 FT base, internally located with respect to any 3' and 5' ends of the DNA
 FT
 PS Example 11; Page 38; 51pp; English.
 XX
 CC The invention relates to single stranded and/or looped DNA having a
 CC portion with at least one base, internally located with respect to any 3'
 CC and 5' ends of the DNA, that is unbasepaired with another base in a form
 CC that is capable of being internalised within a target cell, for use in

CC therapy, with cytosstatic and virucide activity. The DNA acts as an
CC apoptosis inducer in cells that lack p53 functionality, useful for
CC killing a cell, preferably a dividing cell where the cell is other than a
CC Saos-2 cell and the DNA is not configured to express the peptide or
CC protein that selectively kills the cell. The DNA is preferably in the
CC form of an AAV or associated with AAV protein which has been treated such
CC that the DNA is no longer capable of replication or expression in cells
CC and is associated with or contained within a vehicle which is associated
CC with one or more viral fibers which facilitate internalisation of the DNA
CC into a target cell. The DNA is also useful for manufacturing a medicament
CC for treating an individual suffering from a mutant p53 associated cancer
CC or an infection that inhibits cellular p53. The method targets cancer
CC cells or cells infected with p53 inhibiting viruses, such as HPV16 or
CC HPV18 and only cells that lack p53 activity are killed and no damage to
CC cellular DNA is involved. The present sequence is that of the
CC adeno-associated virus (AAV) 2 ITR, the double loop structure found at
CC each end of the AAV DNA.

SQ Sequence 145 BP; 24 A; 49 C; 52 G; 20 T; 0 other;

Query Match 75.1%; Score 107.4; DB 24; Length 145;

Best Local Similarity 85.1%; Pred. No. 3e-19; Mismatches 21; Indels 0; Gaps 0;

Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTTAGTATGATGAGTTGCCCACTCCCTCTGCGCGCTGCTGCTGCGTGGGCGCTG 62
Db 5 ACCCTTAGTATGATGAGTTGCCCACTCCCTCTGCGCGCTGCTGCTGCGTGGGCGCTG 64

QY 63 CGGACCAAGGTCGCGACGCGGCAAGCTCTGCTGCTGCGCGCGCGCCACCGGCGAGCGAGC 122

Db 65 GCGACCAAGGTCGCGACGCGCGCGCGCTTGGCCCGGCGCGCTCACTGAGCGAGCGAGC 124

QY 123 GCGGAGAGAGGAGTGGGCA 143

Db 125 GCGGAGAGAGGAGTGGGCA 145

RESULT 9

AA066769 standard; DNA, 165 BP.

XX AA066769;

DT 20-JAN-1995 (first entry)

DE Double-D ITR terminal repeat.

KW Double-D ITR, inverted terminal repeat; D-sequence; AAV;

KM adeno-associated virus; capsid; encapsidation; gene therapy; vector;

XX de.

OS Synthetic.

PN W09413788-A.

PD 23-JUN-1994.

PF 03-DEC-1993; 93WO-US11728.

PR 04-DEC-1992; 92US-0989841.

PA (UYPI-) UNIV PITTSBURGH.

PI Samulski RJ, Xiao X;

DR WPI; 1994-217866/26.

PT Adenovirus associated viral inverted terminal repeat - for use in

PS diseases a recombinant viral vector system for treatment of genetic

CC The 20-bp D-sequence given in AA066773 is present in the inverted
CC terminal repeat (ITR) sequence of AAV and is required for viral
CC replication. A novel, modified terminal repeat structure, double-D
CC ITR, was constructed (AA066769) that contained a single 145 bp ITR
CC sequence with an additional 'D' sequence. The double-D ITR fragment
CC allows replication and encapsidation of recombinant DNA into AAV
CC capsid proteins. Vectors including double-D ITR can be used for
CC gene replacement therapies.

SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;

Query Match 75.1%; Score 107.4; DB 15; Length 165;

Best Local Similarity 85.1%; Pred. No. 3e-19; Mismatches 21; Indels 0; Gaps 0;

Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTTAGTATGATGAGTTGCCCACTCCCTCTGCGCGCTGCTGCTGCGTGGGCGCTG 62
Db 5 ACCCTTAGTATGATGAGTTGCCCACTCCCTCTGCGCGCTGCTGCTGCGTGGGCGCTG 64

QY 63 CGGACCAAGGTCGCGACGCGGCAAGCTCTGCTGCTGCGCGCGCGCCACCGGCGAGCGAGC 122

Db 65 GCGACCAAGGTCGCGACGCGCGCGCGCTTGGCCCGGCGCGCTCACTGAGCGAGCGAGC 124

QY 123 GCGGAGAGAGGAGTGGGCA 143

Db 125 GCGGAGAGAGGAGTGGGCA 145

RESULT 10

AA09462 standard; DNA, 165 BP.

XX AA09462;

DT 10-SEP-1997 (first entry)

DE Adeno associated virus inverted terminal repeat with double D region.

KW Viral replication; REP protein; inverted terminal repeat; ITR;

KM adeno-associated virus; AAV; vector; double-D; cis-acting;

XX lytic life cycle; gene therapy; ss.

OS Adeno associated virus.

OS Synthetic.

FT Key

FT Location/Qualifiers

FT repeat_unit

FT 1..145

FT /*tag= a

FT /rpt_type= INVERTED

FT /standard_name= ITR

FT /note= "This naturally occurring 145 bp ITR is

FT located at both ends of the AAV genome"

FT 1..20

FT /*tag= b

FT /rpt_type= INVERTED

FT /label= D_repeat

FT 21..62

FT /*tag= c

FT /rpt_type= INVERTED

FT /label= A_repeat

FT /note= "Forms the stem of a T-shaped structure when

FT base paired with repeat A"

FT 63..70

FT /*tag= d

FT /rpt_type= INVERTED

FT /label= C_repeat

FT 74..81

FT /*tag= e

FT /rpt_type= INVERTED

FT /label= C_repeat

FT 85..92

FT /*tag= f

FT /rpt_type= INVERTED

```
FT FT /label= B_repeat
FT 96..103
FT /*tag= g
FT /rpt_type= INVERTED
FT /label= B'_repeat
FT 104..145
FT /*tag= h
FT /rpt_type= INVERTED
FT /label= A_repeat
FT /note= "Forms the stem of a T-shaped structure when
FT base paired with A' repeat"
FT 146..165
FT /*tag= 1
FT /label= D'
FT /note= "Additional D' sequence"
FT 61..82
FT /*tag= j
FT /note= "Part of T-shaped structure, formed by base
FT pairing between the C' and C repeats"
FT stem_loop
FT 84..104
FT /*tag= k
FT /note= "Part of T-shaped structure, formed by base
FT pairing between the B' and B repeats"
FT misc_structure
FT 21..145
FT /*tag= 1
FT /label= T-shaped
FT /note= "The naturally occurring 145 bp ITR sequence
FT can form a T-shaped structure for DNA
FT replication when single-stranded"
FT repeat_region
FT 1..165
FT /*tag= m
FT /label= Double-D
FT /note= "This 165 bp sequence, resulting from
FT addition of a second D repeat to the AAV
FT ITR, has not been identified in any
FT naturally occurring virus"
FT MO9636364-A1.
FT 21-NOV-1996.
FT 14-MAY-1996; 96WO-US06786.
FT 15-MAY-1995; 95US-0440738.
FT (SAMU/) SAMULSKI R J.
FT (XIAO/) XIAO X.
FT Samuelski RJ, Xiao X;
FT WPI; 1997-042643/04.
FT Double-D sequence directs adeno-associated virus integration into
FT host genome - used in gene therapy, maintains full length coding
FT sequence of therapeutic gene
FT Claim 1; Fig 9; 55pp; English.
XX The adeno associated virus (AAV) has a 145 bp inverted terminal
CC repeat (ITR) located at each end of its genome. In addition to
CC being able to base pair with each other, the ITRs can also
CC individually fold back on themselves through the base pairing of A,
CC A', B, B' and C, C' sequences to form a T-shaped structure for DNA
CC replication (see features table). It has been found that viral
CC mutants with deleted D sequences are unable to replicate their DNA.
CC Addition of a second D repeat (D') to the naturally occurring ITR
CC resulted in a sequence, designated double-D, which was sufficient
CC to carry out the functions normally required of two wild-type ITRs
CC during a lytic AAV viral infection, i.e. it is capable of directing
CC replication and assembly into AAV, and/or the integration into the
CC host genome, of recombinant DNA containing the nucleic acid molecule.
CC Vectors and viral particles containing the double-D sequence are
CC useful in gene therapy. Replication and integration into the host
```

```
CC genome is completely effected through the double-D sequences, ensuring
CC that the heterologous gene sequences remain intact.
XX Sequence 165 BP, 27 A, 56 C, 56 G, 26 T, 0 other;
SQ Query Match 75.1%; Score 107.4; DB 18; Length 165;
Best Local Similarity 85.1%; Pred. No. 3e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTGATGATGAGTGGCCCACTCCCTCTGCGGCTTCGCTCGGTGGGCGCTG 62
DB 5 ACCCTTGATGATGAGTGGCCCACTCCCTCTGCGGCTTCGCTCGGTGGGCGCTG 64
QY 63 CGGACCAAGAGTCCGAGAGCGGAGAGCTGCTGCTGCTGCGGCGGACCGAGCGAGC 122
DB 65 GCGACCAAGAGTCCGAGAGCGGAGAGCTGCTGCTGCTGCGGCGGACCGAGCGAGC 124
QY 123 GCGCAGAGAGGAGTGGGCAA 143
DB 125 GCGCAGAGAGGAGTGGGCAA 145
RESULT 11
AAD03535
ID AAD03535 standard; DNA; 165 BP.
XX AAD03535;
AC 19-JUN-2001 (first entry)
DT Inverted terminal repeat double DD DNA sequence.
DE
XX Inverted terminal repeat double DD DNA sequence.
KW Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;
KW inverted terminal repeat; haemostatic; antisticking; neuroprotective;
KW antianaemic; noctropic; blood disorder; sickle cell anaemia;
KW thalassaemia; neurological disorder; haemophilia; Alzheimer's disease;
KW muscle disorder; Parkinson's disease; gene delivery; erythropoietin;
KW epo; CFTR; cystic fibrosis transmembrane conductance receptor;
KW tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;
KW gene therapy; Double DD; ds.
XX
XX Unidentified.
OS
XX
XX MO200125465-A1.
PN 12-Apr-2001.
XX
XX 06-OCT-2000; 2000WO-US27863.
PF
XX 07-OCT-1999; 99US-0158209.
PR
XX (TOMA ) UNIV IOWA RES FOUND.
PA (ENGELHARDT J F.
PA (DONG/) DONGSHENG D.
PA (ZIYI/) ZIYING Y.
XX
XX Engelhardt JF, Dongsheng D, Ziyang Y;
XX WPI; 2001-266321/27.
XX
XX Composition for transferring recombinant DNAs and to express a
XX polypeptide in a host cell, comprises two recombinant adeno-associated
XX viruses -
XX Disclosure; Page 139; 144pp; English.
XX The patent discloses a composition comprising at least two recombinant
XX adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA
XX molecule comprising three DNA segments linked together. The first and
XX the third DNA segments comprise 5' and 3' inverted terminal repeats
XX (ITRs) respectively from the circular intermediate of AAV. The second
XX DNA segment in each virus is different and does not comprise AAV
XX sequence. It preferably comprises sequences encoding a therapeutically
```

CC effective polypeptide such as the cystic fibrosis transmembrane
CC conductance receptor gene (CFTR), the erythropoietin (epo) gene, the
CC tyrosine hydroxylase gene (TH), the erythropoietin (epo) gene, the
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV
CC imparts increased episomal stability and persistence of the vector in
CC the host cell. Compositions comprising AAV sequences are useful for
CC transferring recombinant DNAs to a host cell and express a polypeptide
CC in a host cell. The recombinant vector is useful in medical therapy,
CC cell anaemia, thalassemia, haemophilia, neurological disorders, such
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery
CC vehicle for gene therapy.
CC The present sequence is the ITR DD DNA sequence referred to as
CC "double sequence". This sequence is disclosed in U.S. Patent No.
CC 5,478,745. This sequence is not used in the present invention.
XX
SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;
Query Match 75.1%; Score 107.4; DB 22; Length 165;
Best Local Similarity 85.1%; Pred. No. 3e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTAGTATGATGAGTTGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCGGCGCTG 62
DB 5 ACCCTTAGTATGATGAGTTGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCTGCGGCGCTG 64
QY 63 CGGACCAAAGTCCGACAGCGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
DB 65 GCGACCAAAGTCCGACAGCGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 124
QY 123 GCGCAGAGAGGAGTGGGCAA 143
DB 125 GCGCAGAGAGGAGTGGGCAA 145
RESULT 12
AA03384
ID AA03384 standard; cDNA; 192 BP.
XX
AC AA03384;
XX
DT 18-APR-1996 (first entry)
XX
DE Right-hand inverted terminal repeat from AAV-2, used in pXL2384.
XX
KW inverted terminal repeat; ITR; adenovirus; adeno-associated virus;
XX AAV-2; replication defective; integration; gene therapy; ds.
XX
OS Adeno associated virus 2.
XX
XX W09523867-A1.
XX
XX 08-SEP-1995.
XX
XX 28-FEB-1995; 95MO-FR00233.
XX
XX 03-MAR-1994; 94FR-0002445.
XX
XX (RHON) RHONE POULENC RORER SA.
XX
XX Deneffe P, Latra M, Perricaudet M, Vigne B;
XX WPI; 1995-320581/41.
XX
XX Recombinant defective adenovirus conf. integratable expression
XX cassette - for use in gene therapy to express protein, antigen or
XX anti-sense nucleic acid, also for prodn. of recombinant
XX adeno-associated viruses
XX
XX Example 2; Page 26; 50pp; French.
XX
XX Recombinant, nonpathogenic adenovirus which are able to integrate

CC stably into a host genome are claimed. The viruses pref. contain at
CC least one inverted terminal repeat (ITR) sequence and in particular
CC two ITRs flank a heterologous DNA insert. The present sequence is
CC that of the right-hand ITR from adeno associated virus AAV-2 which
CC was used in construct pXL2384; in pXL2384 the beta-galactosidase
CC gene is flanked by AAV-2 left and right ITRs. The defective viruses
CC are useful for stably introducing large fragments of heterologous DNA
CC making them suitable for gene therapy.
XX
SQ Sequence 192 BP; 39 A; 58 C; 63 G; 32 T; 0 other;
Query Match 75.1%; Score 107.4; DB 16; Length 192;
Best Local Similarity 85.1%; Pred. No. 3.1e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTAGTATGATGAGTTGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCTGCTGCTG 62
DB 52 ACCCTTAGTATGATGAGTTGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCTGCTGCTG 111
QY 63 CGGACCAAAGTCCGACAGCGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
DB 112 GCGACCAAAGTCCGACAGCGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171
QY 123 GCGCAGAGAGGAGTGGGCAA 143
DB 172 GCGCAGAGAGGAGTGGGCAA 192
RESULT 13
AA247166
ID AA247166 standard; DNA; 272 BP.
XX
AC AA247166;
XX
DT 28-MAR-2000 (first entry)
XX
DE Head-to-tail ITR sequence from recombinant AAV vector p1202.
XX
XX Antisickling; hemostatic; neuroprotective; antiparkinsonian; ITR;
XX gene therapy; inverted terminal repeat; vector; episomal stability;
XX blood; neurological; muscle; disorder; ds.
XX
OS Adeno associated virus.
XX
XX W09960146-A1.
XX
XX 25-NOV-1999.
XX
XX 20-MAY-1999; 99MO-US11197.
XX
XX 20-MAY-1998; 98US-0086166.
XX
XX 25-MAR-1999; 99US-0276625.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX (ENGELHARDT J F.
XX (DUAN/) DUAN D.
XX
XX Engelhardt JF, Duan D;
XX WPI; 2000-062462/05.
XX
XX Recombinant adeno-associated virus vector useful for gene therapy
XX against disorders related to blood, neurological and muscular systems
XX
XX Disclosure; Fig 10C; 121pp; English.
XX
XX This sequence corresponds to a head-to-tail inverted terminal repeat
XX (ITR) sequence from the recombinant adeno-associated virus (rAAV)
XX vector p1202. The invention relates to a DNA molecule comprising a DNA
XX segment (or biologically active subunit or variant) of a circular
XX intermediate of an AAV that confers increased episomal stability,
XX #persistence or abundance in a host cell. Compositions containing a

vector with therapeutic gene and delivery vehicle or containing two
vectors expressing a full length polypeptide coordinately, are useful
for manufacturing a medicament for treating pathological conditions or
symptoms in a mammal. The vector is useful for therapeutic or
prophylactic treatments of blood disorders (e.g. sickle cell anemia,
thalassaemias, hemophilias and Fanconi's anemia), neurological disorders
(e.g. Alzheimer's disease, Parkinson's disease) and muscle disorders.

Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;

Query Match 75.1%; Score 107.4; DB 21; Length 272;
Best Local Similarity 85.1%; Pred. No. 3.1e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

3 ACCCTAGTATGATGGAGTTGCCACCTCCTCTGCGGCGCTCGTGGTGGGCGCTG 62
53 ACCCTAGTATGATGGAGTTGGCCACCTCCTCTGCGGCGCTCGTGGTGGGCGCG 112
63 CGGACCAAGGTCGCGACGCGGACGCTGCTGCGGCGCCGACGCGAGCGAGC 122
113 GCGACCAAGGTCGCGCGACGCGGCGCTTGTGCGCGCTCGTGGTGGGCGAGC 172
123 GCGCAGAGAGGAGTGGGCAA 143
173 GCGCAGAGAGGAGTGGGCAA 193

RESULT 14
AAD03534
AAD03534 standard; DNA; 272 BP.

AAD03534;
19-JUN-2001 (first entry)

AAV circular intermediate DNA segment comprising ITRs from clone p1202.

Recombinant adeno-associated virus; RAAV; circular intermediate; ITR;
inverted terminal repeat; haemostatic; antithrombotic; neuroprotective;
antihaemic; noctropic; blood disorder; sickle cell anaemia;
thalassaemia; neurological disorder; haemophilia; Alzheimer's disease;
muscle disorder; Parkinson's disease; gene delivery; erythropoietin;
epo; CPTF; cystic fibrosis transmembrane conductance receptor;
tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;
gene therapy; clone p1202; ds.

Adeno associated virus.
WO200125465-A1.
12-APR-2001.
06-OCT-2000; 2000WO-US27863.
07-OCT-1999; 99US-0158209.
(IOMA) UNIV IOMA RES FOUNO.
(ENGELHARDT J F.
(DONG// DONGSHENG D.
(ZIYI// ZIYING Y.
Engelhardt JF, Dongsheng D, Ziyang Y;
WPI; 2001-266321/27.
Composition for transferring recombinant DNAs and to express a
polypeptide in a host cell, comprises two recombinant adeno-associated
viruses -
Disclosure; Fig 10C; 144pp; English.
The patent discloses a composition comprising at least two recombinant
adeno-associated viruses (RAAVs). Each RAAV contains a recombinant DNA

molecule comprising three DNA segments linked together. The first and
the third DNA segments comprise 5' and 3' inverted terminal repeats
(ITRs) respectively from the circular intermediate of AAV. The second
DNA segment in each virus is different and does not comprise AAV
sequence. It preferably comprises sequences encoding a therapeutically
effective polypeptide such as the cystic fibrosis transmembrane
conductance receptor gene (CFTR), the erythropoietin (epo) gene, the
tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-
brosidase gene (Gaucher's disease). The circular intermediate of AAV
impacts increased episomal stability and persistence of the vector in
the host cell. Compositions comprising RAAV sequences are useful for
transferring recombinant DNAs to a host cell and express a polypeptide
in a host cell. The recombinant vector is useful in medical therapy,
which includes treatment or prophylaxis of blood disorders (e.g. sickle
cell anaemia, thalassaemia, haemophilia), neurological disorders, such
as Alzheimer's disease, Parkinson's disease, muscle disorders involving
skeletal, cardiac or smooth muscle. AAV vector is used as a delivery
vehicle for gene therapy.
The present sequence is circular intermediate DNA segment comprising
ITRs from Adeno associated virus (AAV) clone p1202.

Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 272;
Best Local Similarity 85.1%; Pred. No. 3.1e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

3 ACCCTAGTATGATGGAGTTGCCACCTCCTCTGCGGCGCTCGTGGTGGGCGCTG 62
53 ACCCTAGTATGATGGAGTTGGCCACCTCCTCTGCGGCGCTCGTGGTGGGCGCG 112
63 CGGACCAAGGTCGCGACGCGGACGCTGCTGCGGCGCCGACGCGAGCGAGC 122
113 GCGACCAAGGTCGCGCGACGCGGCGCTTGTGCGCGCTCGTGGTGGGCGAGC 172
123 GCGCAGAGAGGAGTGGGCAA 143
173 GCGCAGAGAGGAGTGGGCAA 193

RESULT 15
AAH41481
AAH41481 standard; DNA; 4675 BP.

AAH41481;
23-AUG-2001 (first entry)

Adeno-associated virus (AAV) plasmid pAV1 Avari nucleotide sequence.

Adeno-associated virus; AAV; gene transfer; adenovirus; chromosome 19;
medicine; cell engineering; genetic engineering; treatment;
embryological engineering; ds.

Adeno-associated virus.
WO200132899-A1.
10-MAY-2001.
23-OCT-2000; 2000WO-JP07373.
29-OCT-1999; 99JP-0308839;
(TAKI) TAKARA SHUZO CO LTD.
Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K;
Asada K, Kato I;
WPI; 2001-316450/33.
Gene transfer method for highly efficient introduction of foreign genes
to cells in humans particularly by integration specifically into AAVS1

PT site of chromosome 19 for transformation, applicable e.g. in disease
PT treatment -

XX
PS Example 1; Page 28-31; 38pp; Japanese.

CC The present invention describes a gene transfer method for highly
CC efficient introduction of foreign genes to cells in humans particularly
CC by integration specifically into AAVS1 site of chromosome 19 for
CC transformation, applicable e.g. in disease treatment. The method
CC comprises transferring into the cells (with the use of an adenovirus
CC vector) a nucleic acid which has a sequence provided with adeno-
CC associated virus (AAV)-originated inverted terminal repeats (ITRs) in
CC both sides of the target foreign gene to be transferred, a second nucleic
CC acid which has an AAV-originated rep gene and a promoter for expressing
CC this gene and carries a stuffer sequence inserted into it sandwiched in
CC two recombinase-recognition sequences and located between the rep gene
CC and promoter and expressing the Rep protein under the action of the
CC recombinase in the cells obtained previously to integrate the target
CC foreign gene into the chromosomal DNA. The method is for transferring a
CC foreign gene into cells particularly in human, especially by integration
CC specifically into AAVS1 site of chromosome 19 for transformation, which
CC is applicable in medicine, as well as cell, genetic and embryological
CC engineering e.g. in disease treatment. The method is highly efficient,
CC with use of adenovirus vectors, nucleic acids and other sequences
CC including rep genes and promoters, by expressing rep proteins to
CC integrate target/foreign gene into chromosomal DNA through action of
CC recombinase. The present sequence represents an AAV plasmid pAV1 Avali
CC nucleotide sequence which is used in an example from the present
CC invention.

XX
SQ Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 4675;

Best Local Similarity 85.1%; Pred. No. 3.5e-19; Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY	3	ACCCCTAGTATGATGAGTTCCTCCCTCTCTGCGGCTCGCTCGCTCGGATGAGGCTG	62
DB	4535	ACCCCTAGTATGATGAGTTCCTCCCTCTCTGCGGCTCGCTCGCTCGGATGAGGCTG	4594
QY	63	CGGACCAAGGTCGCGAGACGCGAGAGCTCTGCTTCGCGGCCCGACCGAGCGAGC	122
DB	4595	GCGACCAAGGTCGCGCGAGCGCGGCTTTCGCGGCCCGCTCACTGAGCGAGC	4654
QY	123	GCGCAGAGGAGGAGTGGCAA	143
DB	4655	GCGCAGAGGAGGAGTGGCAA	4675

Search completed: July 20, 2003, 09:39:48
Job time: 33.1207 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 08:54:01 ; Search time 6.68705 Seconds

(without alignments)
6558.165 Million cell updates/sec

Title: US-09-807-802A-1_COPY_4576_4718

Perfect score: 143

Sequence: 1 ttaccctagtgatgagatc.....cgcagagagggagtgagca 143

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents NA.*
2: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES:

Result No.	Score	Query Match	Length	ID	Description
C 1	107.4	75.1	145	4	US-08-525-866-1 Sequence 1, Appl
2	107.4	75.1	165	1	US-07-989-841A-1 Sequence 1, Appl
3	107.4	75.1	165	2	US-08-440-738A-1 Sequence 1, Appl
4	107.4	75.1	165	3	US-08-471-914-1 Sequence 1, Appl
5	107.4	75.1	165	4	US-08-276-625-7 Sequence 7, Appl
6	107.4	75.1	192	3	US-08-702-573-3 Sequence 3, Appl
7	107.4	75.1	272	4	US-08-254-358-1 Sequence 6, Appl
8	107.4	75.1	4680	1	US-08-254-358-1 Sequence 1, Appl
9	107.4	75.1	4680	2	US-08-475-391-1 Sequence 1, Appl
10	107.4	75.1	4680	1	US-08-709-609-1 Sequence 1, Appl
11	107.4	75.1	4680	5	PCT-US95-07178-1 Sequence 1, Appl
12	107.4	75.1	8698	4	US-09-770-315-2 Sequence 2, Appl
13	105.8	74.0	272	4	US-09-276-625-4 Sequence 4, Appl
C 14	105.4	73.7	174	4	US-09-394-110A-1 Sequence 1, Appl
C 15	102.6	71.7	272	4	US-09-276-625-6 Sequence 6, Appl
C 16	101	70.6	183	4	US-09-394-110A-2 Sequence 2, Appl
C 17	99.4	69.5	145	3	US-07-789-917A-1 Sequence 1, Appl
C 18	99.4	69.5	145	3	US-08-702-573-4 Sequence 4, Appl
C 19	99.4	69.5	145	4	US-07-982-193-1 Sequence 1, Appl
C 20	99.4	69.5	165	1	US-07-989-841A-1 Sequence 1, Appl
C 21	99.4	69.5	165	2	US-08-440-738A-1 Sequence 1, Appl
C 22	99.4	69.5	165	3	US-08-471-914-1 Sequence 1, Appl
C 23	99.4	69.5	165	4	US-08-276-625-7 Sequence 7, Appl
C 24	99.4	69.5	4680	1	US-08-254-358-1 Sequence 1, Appl
C 25	99.4	69.5	4680	1	US-08-475-391-1 Sequence 1, Appl
C 26	99.4	69.5	4680	1	US-08-709-609-1 Sequence 1, Appl
C 27	99.4	69.5	4680	5	PCT-US95-07178-1 Sequence 1, Appl

28	99.4	69.5	5932	4	US-09-299-141-4 Sequence 4, Appl
C 29	99.4	69.5	5932	4	US-09-299-141-4 Sequence 4, Appl
30	99.4	69.5	6142	4	US-09-299-141-8 Sequence 8, Appl
C 31	99.4	69.5	6142	4	US-09-299-141-8 Sequence 8, Appl
32	99.4	69.5	6253	3	US-08-893-327-15 Sequence 15, Appl
C 33	99.4	69.5	6253	3	US-08-893-327-15 Sequence 15, Appl
34	99.4	69.5	6280	3	US-08-893-327-17 Sequence 17, Appl
C 35	99.4	69.5	6280	3	US-08-893-327-17 Sequence 17, Appl
36	99.4	69.5	6280	3	US-08-893-327-19 Sequence 19, Appl
C 37	99.4	69.5	6280	3	US-08-893-327-19 Sequence 19, Appl
38	99.4	69.5	6565	4	US-09-299-141-1 Sequence 1, Appl
C 39	99.4	69.5	6565	4	US-09-299-141-1 Sequence 1, Appl
40	99.4	69.5	6714	4	US-09-299-141-6 Sequence 6, Appl
C 41	99.4	69.5	6714	4	US-09-299-141-6 Sequence 6, Appl
42	99.4	69.5	6924	4	US-09-299-141-9 Sequence 9, Appl
C 43	99.4	69.5	6924	4	US-09-299-141-9 Sequence 9, Appl
44	99.4	69.5	6924	4	US-09-299-141-10 Sequence 10, Appl
C 45	99.4	69.5	6924	4	US-09-299-141-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-525-866-1/c
Sequence 1, Application us/08525866
Patent No. 6207457
GENERAL INFORMATION:
APPLICANT: NATSOLIS, GEORGES
APPLICANT: FUROSKY, RICHARD T.
TITLE OF INVENTION: TARGETED NUCLEOTIDE SEQUENCE DELIVERY
TITLE OF INVENTION: AND INTEGRATION SYSTEM
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: REED & ROBINS
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,866
FILING DATE: 08-SEP-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: ROBERTS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 0800-0006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-525-866-1
Query Match 75.1%; Score 107.4; DB 4; Length 145;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCCAGTATGAGATGCGCACTCCCTCGCGCGCTCGCTCGTGGGCGCTG
Db 141 ACCCTTAGTATGAGATGCGCACTCCCTCGCGCGCTCGCTCGTCAAGGCGG 82
QY 63 CGACCAAGGTCCGACGAGAGAGAGCTGCTGCGCGCGCCCAACCGAGCGAGC 122

Db 81 GCGACCAAGGTCCGCCACCCCGGCTTTGGCCCGGCTCTAGAGAGCGAGC 22
QY 123 GCGCAGAGAGGAGTGGCCAA 143
Db 21 GCGCAGAGAGGAGTGGCCAA 1

RESULT 2

US-07-989-841A-1
Sequence 1, Application US/07989841A
Patent No. 5478745
GENERAL INFORMATION:
APPLICANT: Samulski, R. J.
TITLE OF INVENTION: Recombinant Viral Vector System
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/989, 841A
FILING DATE: On even date herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6636-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-07-989-841A-1

Query Match 75.1%; Score 107.4; DB 1; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCGCTCTGCTGCTGAGGCGCTG 62
Db 5 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCGCTCTGCTGCTGAGGCGG 64
QY 63 CGACCAAGGTCCGCGAGAGCGGAGAGTCTGCTCTGCGGCGCCCAAGAGCGAGCG 122
Db 65 GCGACCAAGGTCCGCGAGAGCGGAGAGTCTGCTCTGCGGCGGCGCTCTGAGAGCGAGC 124
QY 123 GCGCAGAGAGGAGTGGCCAA 143
Db 125 GCGCAGAGAGGAGTGGCCAA 145

RESULT 3

US-08-440-738A-1
Sequence 1, Application US/08440738A
Patent No. 5869305
GENERAL INFORMATION:
APPLICANT: Samulski, R. J.

APPLICANT: Xiao, X.
TITLE OF INVENTION: Recombinant Viral Vector System
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440, 738A
FILING DATE: May 15, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6636-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-440-738A-1

Query Match 75.1%; Score 107.4; DB 2; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCGCTCTGCTGCTGAGGCGCTG 62
Db 5 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCGCTCTGCTGCTGAGGCGG 64
QY 63 CGACCAAGGTCCGCGAGAGCGGAGAGTCTGCTCTGCGGCGCCCAAGAGCGAGCG 122
Db 65 GCGACCAAGGTCCGCGAGAGCGGAGAGTCTGCTCTGCGGCGGCGCTCTGAGAGCGAGC 124
QY 123 GCGCAGAGAGGAGTGGCCAA 143
Db 125 GCGCAGAGAGGAGTGGCCAA 145

RESULT 4

US-08-471-914-1
Sequence 1, Application US/08471914A
Patent No. 6057152
GENERAL INFORMATION:
APPLICANT: Samulski, R.
TITLE OF INVENTION: RECOMBINANT VIRAL VECTOR SYSTEM
FILE REFERENCE: 6636-027
CURRENT APPLICATION NUMBER: US/08/471, 914A
CURRENT FILING DATE: 1995-06-06
EARLIER APPLICATION NUMBER: 08/440, 738
EARLIER FILING DATE: 1995-05-15
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 165
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: double-D
OTHER INFORMATION: sequence
US-08-471-914-1

Query Match 75.1%; Score 107.4; DB 3; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 5 ACCCTAGTATGAGTTGGCACTCCCTCTCTGCGGCGCTGCTGCTGCTGAGGCGG 64
QY 63 CGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 65 GCGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 124
QY 123 GCGGAGAGAGGAGTGGGCAA 143
DB 125 GCGGAGAGAGGAGTGGGCAA 145

RESULT 5

US-09-276-625-7
Sequence 7, Application US/09276625
Patent No. 6436392
GENERAL INFORMATION:
APPLICANT: Engelhardt, John F.
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875.007US1
CURRENT APPLICATION NUMBER: US/09/276,625
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/086,166
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 165
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: SEQ ID NO:1 of U.S. Patent No. 6436392 5,478,745
US-09-276-625-7

Query Match 75.1%; Score 107.4; DB 4; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 5 ACCCTAGTATGAGTTGGCACTCCCTCTCTGCGGCGCTGCTGCTGCTGAGGCGG 64
QY 63 CGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 65 GCGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 124
QY 123 GCGGAGAGAGGAGTGGGCAA 143
DB 125 GCGGAGAGAGGAGTGGGCAA 145

RESULT 6

US-08-702-573-3
Sequence 3, Application US/08702573
Patent No. 6033885
GENERAL INFORMATION:
APPLICANT: LATTA, Martine
APPLICANT: DENEFLE, Patrice
TITLE OF INVENTION: INTEGRATIVE RECOMBINANT ADENOVIRUSES,
PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Rd. 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,573
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94/02445
FILING DATE: 03-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/FR95/00233
FILING DATE: 28-FEB-1995

ATTORNEY/AGENT INFORMATION:
NAME: Smith Ph.D., Julie K.
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: ST94011-US
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..192
OTHER INFORMATION: /note= "Right ITR Sequence in
Patent No. 6033885
OTHER INFORMATION: PXL2384"

US-08-702-573-3
Query Match 75.1%; Score 107.4; DB 3; Length 192;
Best Local Similarity 85.1%; Pred. No. 1.8e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 52 ACCCTAGTATGAGTTGGCACTCCCTCTCTGCGGCGCTGCTGCTGCTGAGGCGG 111
QY 63 CGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 122
DB 112 GCGACCAAAAGTCCGCGAGAGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 171
QY 123 GCGGAGAGAGGAGTGGGCAA 143
DB 172 GCGGAGAGAGGAGTGGGCAA 192

RESULT 7

US-09-276-625-6
Sequence 6, Application US/09276625
Patent No. 6436392
GENERAL INFORMATION:
APPLICANT: Engelhardt, John F.
APPLICANT: Duan, Dongsheng
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875.007US1
CURRENT APPLICATION NUMBER: US/09/276,625
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/086,166


```

RESULT 10
US-08-709-609-1
Sequence 1, Application US/08709609
Patent No. 5858775
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709, 609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5858775and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-709-609-1
Query Match 75.1%; Score 107.4; DB 2; Length 4680;
Best Local Similarity 85.1%; Pred. No. 2.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTAGTAGTAGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCGGCGCTG 62
Db 4540 ACCCTAGTAGTAGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTGCTGCGGCGG 4599
QY 63 CGGACCAAAAGTCCGACAGCGGACAGCTTGTCTGTCCCGGCCACCAAGACGACGAC 122
Db 4600 GCGACCAAAAGTCCGACAGCGGCGGCGCTTGTCCCGGCGGCGCTCAGTAGACGACGAC 4659
QY 123 GCGCAGAGAGGAGTGGGCAA 143
Db 4660 GCGCAGAGAGGAGTGGGCAA 4680
RESULT 11
PCT-US95-07178-1
Sequence 1, Application PC/TUS9507178
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois

```

```

1 COUNTRY: USA
2 ZIP: 60606
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patent Release #1.0, Version #1.25
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: PCT/US95/07178
10 FILING DATE:
11 CLASSIFICATION:
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Noland, Greta E.
14 REGISTRATION NUMBER: 35,302
15 REFERENCE/DOCKET NUMBER: 31975
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (312) 474-6300
18 TELEFAX: (312) 474-0448
19 TELEX: 25-3856
20 INFORMATION FOR SEQ ID NO: 1:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 4680 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: single
25 TOPOLOGY: linear
26 MOLECULE TYPE: DNA (genomic)
27 PCT-US95-07178-1

```

Query Match	75.1%	Score 107.4;	DB 5;	Length 4680;
Best Local Similarity	85.1%;	Pred. No. 2.5e-21;		
Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 12
US-09-770-315-2
; Sequence 2, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; PRIOR FILING DATE: 2001-01-26
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 8698
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-2

```

	Query Match	75.1%	Score 107.4	DB 4	Length 8698
	Best Local Similarity	85.1%	Pred. No. 2.7e-21		
	Matches 120	Conservative 0	Mismatches 21	Indels 0	Gaps 0
QY	3 ACCCTGATGATGAGTTGCCCACTCCCTCTCTGCGCGCTGCGTCGCTGGGGCCTG	62			

Db 4539 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 4598
Qy 63 CGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 122
Db 4599 GCGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 4658
Qy 123 GCGCAGAGAGGAGTGGGCA 143
Db 4659 GCGCAGAGAGGAGTGGGCA 4679

RESULT 13
US-09-276-625-4
Sequence 4, Application US/09276625
Patent No. 6436392
GENERAL INFORMATION:
APPLICANT: Engelhardt, John F.
APPLICANT: Duan, Dongsheng
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875,007US1
CURRENT APPLICATION NUMBER: US/09/276,625
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/086,166
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 272
TYPE: DNA
ORGANISM: AAV circular intermediate, clone p81
US-09-276-625-4

Query Match 74.0%; Score 105.8; DB 4; Length 272;
Best Local Similarity 84.4%; Pred. No. 5.2e-21;
Matches 119; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 3 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 62
Db 53 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 112
Qy 63 CGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 122
Db 113 GCGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 172
Qy 123 GCGCAGAGAGGAGTGGGCA 143
Db 173 GCGCAGAGAGGAGTGGGCA 193

RESULT 14
US-09-394-110A-1/c
Sequence 1, Application US/09394110A
Patent No. 6451594
GENERAL INFORMATION:
APPLICANT: Chien, Kenneth
APPLICANT: Wang, Yibin
APPLICANT: Evans, Sylvia
TITLE OF INVENTION: No. 6451594e1 Recombinant Adenovirus for Tissue Specific Expression
FILE REFERENCE: 6627-P8045
CURRENT APPLICATION NUMBER: US/09/394,110A
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.0
SEQ ID NO 1
LENGTH: 174
TYPE: DNA
ORGANISM: adeno-associated virus 2
US-09-394-110A-1

Query Match 73.7%; Score 105.4; DB 4; Length 174;
Best Local Similarity 84.9%; Pred. No. 6.5e-21;
Matches 118; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 62
Db 139 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 80
Qy 63 CGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 122
Db 79 GCGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 20
Qy 123 GCGCAGAGAGGAGTGGGCA 141
Db 19 GCGCAGAGAGGAGTGGGCA 1

RESULT 15
US-09-276-625-6/c
Sequence 6, Application US/09276625
Patent No. 6436392
GENERAL INFORMATION:
APPLICANT: Engelhardt, John F.
APPLICANT: Duan, Dongsheng
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875,007US1
CURRENT APPLICATION NUMBER: US/09/276,625
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/086,166
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 272
TYPE: DNA
ORGANISM: AAV circular intermediate, clone p1202
US-09-276-625-6

Query Match 71.7%; Score 102.6; DB 4; Length 272;
Best Local Similarity 83.0%; Pred. No. 4e-20;
Matches 117; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 3 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 62
Db 209 ACCCTAGTAGTAGGAGTGGCCACTCCCTCTCTGCGCGCTGCTGCTCACTAGAGGCCG 150
Qy 63 CGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 122
Db 149 GCGACCAAAAGGTCCGCAAGCGGAGAGTCTGCTGCGCGCCCGACCGAGCGAGC 90
Qy 123 GCGCAGAGAGGAGTGGGCA 143
Db 89 GCGCAGAGAGGAGTGGGCA 69

Search completed: July 20, 2003, 13:23:49
Job time : 7.68705 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 20, 2003, 13:19:56 ; Search time 30.7204 Seconds
(without alignments)
9603.029 Million cell updates/sec

Title: US-09-807-802a-1_COPY_4576_4718

Perfect score: 143

Sequence: 1 ttaccctagtgatgaggtc.....cgcagagagagtgagtgcaaa 143

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PTC NEW PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107.4	75.1	145	US-10-240-198-2	Sequence 2, Appl1
2	107.4	75.1	165	US-10-054-665-7	Sequence 7, Appl1
3	107.4	75.1	207	US-10-023-208-58	Sequence 58, Appl1
4	107.4	75.1	272	US-10-054-665-6	Sequence 6, Appl1
5	107.4	75.1	4675	US-09-782-378A-1	Sequence 1, Appl1
6	107.4	75.1	4675	US-09-782-378A-2	Sequence 2, Appl1
7	107.4	75.1	4675	US-10-240-198-1	Sequence 1, Appl1
8	107.4	75.1	4679	US-09-804-898-1	Sequence 1, Appl1
9	107.4	75.1	4679	US-09-804-898-10	Sequence 10, Appl1
10	107.4	75.1	4679	US-10-038-972A-12	Sequence 12, Appl1
11	107.4	75.1	4680	US-10-077-294-1	Sequence 1, Appl1
12	107.4	75.1	4680	US-10-163-886-1	Sequence 1, Appl1
13	107.4	75.1	4680	US-10-263-127-1	Sequence 1, Appl1
14	105.8	74.0	272	US-10-054-665-4	Sequence 4, Appl1
15	102.6	71.7	272	US-10-054-665-6	Sequence 6, Appl1
C	99.4	69.5	145	US-09-782-378A-6	Sequence 6, Appl1

C 17	99.4	69.5	146	US-10-135-984-8	Sequence 8, Appl1
C 18	99.4	69.5	165	US-09-782-378A-8	Sequence 8, Appl1
C 19	99.4	69.5	165	US-10-054-665-7	Sequence 7, Appl1
C 20	99.4	69.5	207	US-10-023-208-58	Sequence 58, Appl1
C 21	99.4	69.5	4675	US-09-782-378A-1	Sequence 1, Appl1
C 22	99.4	69.5	4675	US-09-782-378A-2	Sequence 2, Appl1
C 23	99.4	69.5	4675	US-10-240-198-1	Sequence 1, Appl1
C 24	99.4	69.5	4679	US-09-804-898-1	Sequence 1, Appl1
C 25	99.4	69.5	4679	US-09-804-898-10	Sequence 10, Appl1
C 26	99.4	69.5	4679	US-10-038-972A-12	Sequence 12, Appl1
C 27	99.4	69.5	4680	US-10-077-294-1	Sequence 1, Appl1
C 28	99.4	69.5	4680	US-10-163-886-1	Sequence 1, Appl1
C 29	99.4	69.5	4680	US-10-263-127-1	Sequence 1, Appl1
C 30	99.4	69.5	5932	US-10-267-117-4	Sequence 4, Appl1
C 31	99.4	69.5	5932	US-10-267-117-4	Sequence 4, Appl1
C 32	99.4	69.5	5932	US-10-340-112-4	Sequence 4, Appl1
C 33	99.4	69.5	5932	US-10-340-112-4	Sequence 4, Appl1
C 34	99.4	69.5	6142	US-10-267-117-8	Sequence 8, Appl1
C 35	99.4	69.5	6142	US-10-267-117-8	Sequence 8, Appl1
C 36	99.4	69.5	6142	US-10-340-112-8	Sequence 8, Appl1
C 37	99.4	69.5	6142	US-10-340-112-8	Sequence 8, Appl1
C 38	99.4	69.5	6565	US-10-267-117-1	Sequence 1, Appl1
C 39	99.4	69.5	6565	US-10-267-117-1	Sequence 1, Appl1
C 40	99.4	69.5	6565	US-10-340-112-1	Sequence 1, Appl1
C 41	99.4	69.5	6565	US-10-340-112-1	Sequence 1, Appl1
C 42	99.4	69.5	6714	US-10-267-117-6	Sequence 6, Appl1
C 43	99.4	69.5	6714	US-10-267-117-6	Sequence 6, Appl1
C 44	99.4	69.5	6714	US-10-340-112-6	Sequence 6, Appl1
C 45	99.4	69.5	6714	US-10-340-112-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-10-240-198-2
Sequence 2, Application US/10240198
Publication No. US20030100115A1
GENERAL INFORMATION:
APPLICANT: BTG International Ltd
APPLICANT: BEARD DR, PETER
APPLICANT: RAJ DR, KENNETH
TITLE OF INVENTION: CYTOTOXIC AGENTS
FILE REFERENCE: 142184WO
CURRENT APPLICATION NUMBER: US/10/240,198
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: 0009887.1
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 145
TYPE: DNA
ORGANISM: adeno-associated virus 2
FEATURE:
NAME/KEY: misc structure
LOCATION: (1)-(145)
OTHER INFORMATION: ITR
FEATURE:
NAME/KEY: misc feature
LOCATION: (72)
OTHER INFORMATION: Unpaired base
FEATURE:
NAME/KEY: misc feature
LOCATION: (94)
OTHER INFORMATION: Unpaired base
US-10-240-198-2

Query Match 75.1%; Score 107.4; DB 15; Length 145;
Best Local Similarity 85.1%; Pred. No. 2.6e-21;
Matches 120; Conservative 21; Indels 0; Gaps 0;
3 ACCCTAGTATGATGAGTCCACCTCCCTCTGCGCGCTCGCTCGTGGAGGCTTG 62

Db 5 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTCACTGAGGCGG 64
Qy 63 CGGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCCCGACCGAGCGAGC 122
Db 65 GCGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCGGCTCTCACTGAGCGAGC 124
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGGCAA 145

RESULT 2

US-10-054-665-7
Sequence 7, Application US/10054665
Publication No. US20020197237A1
GENERAL INFORMATION:
APPLICANT: Duan, Dongsheng
APPLICANT: Engelhardt, John F.
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875.007US2
CURRENT FILING DATE: 2002-06-13
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 09/276,625
PRIOR FILING DATE: 1998-05-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 165
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: SEQ ID NO:1 of U.S. Patent No. US20020197237A1 5,478,745
US-10-054-665-7

Query Match 75.1%; Score 107.4; DB 15; Length 165;
Best Local Similarity 85.1%; Pred. No. 2,5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTGAGGCGCTG 62
Db 5 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGAGGCGG 64
Qy 63 CGGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCCCGACCGAGCGAGC 122
Db 65 GCGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCGGCTCTCACTGAGCGAGC 124
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGGCAA 145

RESULT 3

US-10-023-208-58
Sequence 58, Application US/10023208
Publication No. US20030124537A1
GENERAL INFORMATION:
APPLICANT: Li, Min
APPLICANT: Liu, Yuan-Ching
TITLE OF INVENTION: PROCARVOTIC LIBRARIES AND USES
FILE REFERENCE: A-70174-1/RPT/RMS/RMK
CURRENT FILING DATE: 2001-12-17
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 63
SOFTWARE: Patent version 3.1
SEQ ID NO 58
LENGTH: 207
TYPE: DNA
ORGANISM: Artificial sequence

FEATURE:
OTHER INFORMATION: synthetic enzyme attachment site sequence
US-10-023-208-58

Query Match 75.1%; Score 107.4; DB 15; Length 207;
Best Local Similarity 85.1%; Pred. No. 2,4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTGAGGCGCTG 62
Db 26 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTGAGGCGG 85
Qy 63 CGGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCCCGACCGAGCGAGC 122
Db 86 GCGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCGGCTCTCACTGAGCGAGC 145
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 146 GCGCAGAGAGGAGTGGGCAA 166

RESULT 4

US-10-054-665-6
Sequence 6, Application US/10054665
Publication No. US20020197237A1
GENERAL INFORMATION:
APPLICANT: Duan, Dongsheng
APPLICANT: Engelhardt, John F.
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875.007US2
CURRENT FILING DATE: 2002-06-13
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 09/276,625
PRIOR FILING DATE: 1998-05-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 272
TYPE: DNA
ORGANISM: AAV circular intermediate, clone p1202
US-10-054-665-6

Query Match 75.1%; Score 107.4; DB 15; Length 272;
Best Local Similarity 85.1%; Pred. No. 2,3e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTGAGGCGCTG 62
Db 53 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGGCTCGTCTGCTGAGGCGG 112
Qy 63 CGGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCCCGACCGAGCGAGC 122
Db 113 GCGACCAAGGTCCGCGACAGCGAGCTCTGCTTCCGCGCGGCTCTCACTGAGCGAGC 172
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 173 GCGCAGAGAGGAGTGGGCAA 193

RESULT 5

US-09-782-378A-1
Sequence 1, Application US/09782378A
Patent No. US20020102731A1
GENERAL INFORMATION:
APPLICANT: Hearing, Patrick
APPLICANT: Bahou, Nadie
APPLICANT: Sandlen, Ziv
APPLICANT: Gatenko, Dmitri
TITLE OF INVENTION: Adenoviral Vectors
FILE REFERENCE: STONY-04970
CURRENT APPLICATION NUMBER: US/09/782,378A

```

:
: CURRENT FILING DATE: 2001-02-12
: PRIOR APPLICATION NUMBER: 60/237,747
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 27
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 1
:
: LENGTH: 4675
:
: TYPE: DNA
:
: ORGANISM: Human adeno-associated virus 2
:
: US-09-782-378A-1

```

Query Match 1%	Score 107.4	DB 11	Length 4675
Best Local Similarity	85.1%	Pred. No. 1.4e-21	
Matches 120	Conservative	0	Mismatches 21
			Indels 0
			Gaps 0

QY	3	ACCCCTAGTATGGAGTTGCCACATCCCTCTGTGCGCCTCGCTGCTCGGAGGAGCCTG	62
Db	4535	ACCCCTAGTATGGAGTTGGCCATCTCCCTCTGTGCGCCTCGCTGCTCACTGAGGCCGG	459
QY	63	CGGACCAAAAGGTCCGACAGCGAGGCTGTGCTGCGGCCACCGAGGACGAGC	122
Db	4595	GGGACCAAAAGGTCCGCCGACCGCCCGGGCTTTGCCCGGGCGGCTCAGTGAAGGACGAGC	465
QY	123	GGCGACAGAGGAGAGTGCGCAA	143
Db	4655	GGCGACAGAGGAGAGTGCGCAA	4675

```

RESULT 6
US-09-782-378A-2
Sequence 2, Application US/09782378A
Patent No. US20020102731A1
GENERAL INFORMATION:
APPLICANT: Hearing, Patrick
APPLICANT: Bahou, Nadie
APPLICANT: Sandakou, Ziv
APPLICANT: Gatenko, Dmitri
TITLE OR INVENTION: Adenoviral Vectors
FILE REFERENCE: STONB-04970
CURRENT APPLICATION NUMBER: US/09/782,378A
CURRENT FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/237,747
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 4675
TYPE: DNA
ORGANISM: Human adeno-associated virus 2
US-09-782-378A-2

```

Query Match	75.1%;	Score 107.4;	DB 11;	Length 4675;
-------------	--------	--------------	--------	--------------

Best Local Similarity 85.1%; Pred. NO. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0;
Gaps 0;

Qy	3	ACCCTAGATGATGAGTTGGCCACATCCCTCTCTGGGCGCTCGCTGCTCGGCGGGAGCTG	62
Db	4535	ACCCCTAGATGATGAGTTGGCCACATCCCTCTCTGGGCGCTCGCTGCTCGCTGACTGAGCGCG	4599
Qy	63	CGGACCAAGGTCGCGAGCGGACAGGCTCTGCTCGCGGCGCCACCGAGGCGAGCGAGC	122
Db	4595	GCGACCAAGGTCGCGCGAGCGGCGGCGGCTTTGCGCGGCGGCGCTCAGTGAAGCGAGCGAGC	4655
Qy	123	GCGCAGAGAGGGAGTGCGCAA	143
Db	4655	GCGCAGAGAGGGAGTGCGCAA	4675

RESULT 7
US-10-240-198-1
; Sequence 1, Application US/10240198
; Publication No. US20030100115A1
; GENERAL INFORMATION:

```

?      APPLICANT: BTG International Ltd
?      APPLICANT: BEARD DR, PETER
?      APPLICANT: RAJ DR, KENNETH
?      TITLE OF INVENTION: CYTOTOXIC AGENTS
?      FILE REFERENCE: 142184WO
?      CURRENT APPLICATION NUMBER: US/10/240,198
?      CURRENT FILING DATE: 2002-09-30
?      PRIOR APPLICATION NUMBER: 0009887.1
?      PRIOR FILING DATE: 2000-04-20
?      NUMBER OF SEQ ID NOS: 6
?      SOFTWARE: PatentIn Ver. 2.1
?      SEQ ID NO 1
?      LENGTH: 4675
?      TYPE: DNA
?      ORGANISM: adeno-associated virus 2
US-10-240-198-1

```

Query Match	75.1%	Score 107.4	DB 15	Length 4675
Best Local Similarity	85.1%	Pred. No. 1.4e-21		
Matches 120; Conservative		0; Mismatches 21	Indels 0	Gaps 0

QY	3	ACCCCTAGTGTGTGAGAGTTGGCCACATCCCTCTCTGTGGCGCTGCGTGTCTCGAGAGCCCG	b2
Db	4535	ACCCTAGTGTGTGAGAGTTGGCCACATCCCTCTCTGTGGCGCTGCGTGTCTCGAGAGCCCG	4590
QY	63	CGGACCAAAAGTCCGCGACAGCGAGCTCTGTCTGTGCGGCCACCGAGCGAGCGAC	122
Db	4595	GGAGCAAAAGTCTGGCCGACGCCCGCGGCTTTGGCCCGGCGGCTCATGTGAGGAGCGAGC	465
QY	123	GGCGCAGAGAGGAGTGGGCAA	143
Db	4655	GGCGCAGAGAGGAGTGGCCAA	4675

```

RESULT 8
US-09-804-898-1
; Sequence 1, Application US/09804898
; Patent No. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: DURING, MATTHEW
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804, 898
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189,110
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-804-898-1

```

Query Match	75.1%;	Score 107.4;	DB 10;	Length 4679;
Post local similarity:	85.1%;	Prod No 148-21;		

best local similarity 83.15; file: NO. 1.14-22;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0

QY	3	ACCCTAGTGTGTGAGTTGGCCACATCCCTCTCTGTGGCCTCGCTCGCTCGGTGGGAGCC	62
Db	4539	ACCCTAGTGTGTGAGTTGGCCACATCCCTCTCTGTGGCCTCGCTCGCTACAGAGGCCG	4591
QY	63	CGAGCAAAAGTCCGCGACGCGAGAGCTGTGTGTGGCGGCCCAACGAGGAGCGAGC	122
Db	4559	CGAGCAAAAGTCCGCGCGACGCGCGGTTTGTCCCGGCGGCTCATGTAGAGGAGCGAGC	4655
QY	123	GCGCAGAGAGGGAGTGGGCAA	143
Db	4659	GCGCAGAGAGGGAGTGGGCCAA	4679

RESULT 5

```

US-09-945-681-10
; Sequence 161, Application US/09945681
; Patent No. US20020064878A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITE DE NANTES
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION
; FILE REFERENCE: B4182A - UNIVERSITE DE NANTES
; CURRENT APPLICATION NUMBER: US/09/945,681
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: PCT/EP 00/01854
; PRIOR FILING DATE: 2000-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
; US-09-945-681-10

```

Query Match	75.1%	Score 107.4	DB 10	Length 4679
Best Local Similarity	85.1%	Pred No. 1.4e-21		
Matches 120	Conservative	0	Mismatches 21	Indels 0
			Gaps 0	
QY	3	ACCCCTAGTAGTAGGATGCCACATCCCTCTCTGTGGCGCTGCTGCTGCGTGGGGGCGCTG	62	
Db	4539	ACCCTTAGTAGTAGGATGGGCACATCCCTCTGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4598	
QY	63	CGACCAAAAGTCCGCAAGCGCAGAGCTTGTCTTGCCCGATCCCAACGAGCGAGCGAGC	122	
Db	4599	GCGACCAAAAGTCCGCGCCAGCGCCCGGGCTTTTGGCCGCGCGCGCTCATGAGAGGAGCGAGC	4658	
QY	123	GCGCAGAGGAGGAGTGGGCAA	143	
Db	4659	GCGCAGAGGAGGAGTGGCCAA	4679	

```

RESULT 10
US-10-038-972A-12
; Sequence 12, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-038-972A-12

```

	Query Match	Similarity	Score	ID	Length
Best Local	85.1%	107.4	DB 15	4679	
Matches	120	Conservative	0	Mismatches	21
				Indels	0
				Gaps	0

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/163,886
FILING DATE: 04-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/292,703
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. US20020187129Aland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-163-886-1

Query Match 75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTGAGGCGCTG 62
DB 4540 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTGAGGCGCG 4599

QY 63 CGGACCAAGGTCCGAGAGCGGAGAGCTCTGCTCTGCGCGCCCAACGAGCGAGC 122
DB 4600 GCGACCAAGGTCCGAGAGCGGAGAGCTCTGCTCTGCGCGCCCAACGAGCGAGC 4659

QY 123 GCGCAGAGAGGAGTGGGCAA 143
DB 4660 GCGCAGAGAGGAGTGGGCAA 4680

RESULT 13
US-10-263-127-1
Sequence 1, Application US/10263127
Publication No. US20030082145A1
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/263,127
FILING DATE: 02-Oct-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/10/077,294
FILING DATE: 15-Feb-2002
APPLICATION NUMBER: 09/691,604
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. US20030082145Aland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-263-127-1

Query Match 75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTGAGGCGCTG 62
DB 4540 ACCCTAGTATGATGAGTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTGAGGCGCG 4599

QY 63 CGGACCAAGGTCCGAGAGCGGAGAGCTCTGCTCTGCGCGCCCAACGAGCGAGC 122
DB 4600 GCGACCAAGGTCCGAGAGCGGAGAGCTCTGCTCTGCGCGCCCAACGAGCGAGC 4659

QY 123 GCGCAGAGAGGAGTGGGCAA 143
DB 4660 GCGCAGAGAGGAGTGGGCAA 4680

RESULT 14
US-10-054-665-4
Sequence 4, Application US/10054665
Publication No. US20020197237A1
GENERAL INFORMATION:
APPLICANT: Engelhardt, John F.
TITLE OF INVENTION: Adeno-associated virus vectors
FILE REFERENCE: 875,007US2
CURRENT APPLICATION NUMBER: US/10/054,665
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 09/276,625
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/086,166
PRIOR FILING DATE: 1998-05-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 272
TYPE: DNA
ORGANISM: AAV circular intermediate, clone p81
US-10-054-665-4

Query Match 74.0%; Score 105.8; DB 15; Length 272;
Best Local Similarity 84.4%; Pred. No. 6.4e-21;
Matches 119; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

QY 3 ACCCTAGTAGTGAAGTTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTTG 62
   |||||
Db 53 ACCCTAGTAGTGAAGTTGGCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCCGG 112
   |||||
QY 63 CGACCAAAGGTCGCGACAGCGGAGCTCTGCTCTGCGCGCTCGCGCTCGCGAGCGAGC 122
   |||||
Db 113 GCGGCAAAAGGTCGCGCGAGCGCGCGCTTGCCCGGCGCGCTCAGTAGCGAGCGAGC 172
   |||||
QY 123 GCGGAGAGAGGAGTGGGCAA 143
   |||||
Db 173 GCGGAGAGAGGAGTGGCAA 193
   |||||

```

RESULT 15

```

US-10-054-665-6/C
; Sequence 6, Application US/10054665
; Publication No. US20020197237A1
; GENERAL INFORMATION:
; APPLICANT: Engelhardt, John F.
; APPLICANT: Duan, Dongsheng
; TITLE OF INVENTION: Adeno-associated virus vectors
; FILE REFERENCE: 875.007US2
; CURRENT APPLICATION NUMBER: US/10/054,665
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/276,625
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/086,166
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 272
; TYPE: DNA
; ORGANISM: AAV circular intermediate, clone p1202
US-10-054-665-6

```

```

Query Match 71.7%; Score 102.6; DB 15; Length 272;
Best Local Similarity 83.0%; Pred. No. 5e-20;
Matches 117; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 3 ACCCTAGTAGTGAAGTTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTTG 62
   |||||
Db 209 ACCCTAGTAGTGAAGTTGGCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCCGG 150
   |||||
QY 63 CGACCAAAGGTCGCGACAGCGGAGCTCTGCTCTGCGCGCTCGCGCTCGCGAGCGAGC 122
   |||||
Db 149 GCGACCAAAGGTCGCGCGAGCGCGCTTGCGCGCTCGCGCTCGAGCGAGCGAGC 90
   |||||
QY 123 GCGGAGAGAGGAGTGGGCAA 143
   |||||
Db 89 GCGGAGAGAGGAGTGGCAA 69
   |||||

```

Search completed: July 20, 2003, 20:11:28
 Job time : 30.7204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 377.075 Seconds

(without alignments)
11036.790 Million cell updates/sec

Title: US-09-807-802a-1_COPY_4576_4718

Perfect score: 143

Sequence: 1 ttaccctagtgatgaggtc.....cgacagagaggtgagca 143

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb Da:*
2: gb Htg:*
3: gb In:*
4: gb Cm:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vi:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pat:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*
29: em Vi:*
30: em Htg_Hum:*
31: em Htg_In:*
32: em Htg_Other:*
33: em Htg_Mus:*
34: em Htg_Pln:*
35: em Htg_Rod:*
36: em Htg_Mam:*
37: em Htg_Vit:*
38: em Sv:*
39: em Htgo_Hum:*
40: em Htgo_Mus:*
41: em Htgo_Other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	4718	14 AF063497	AF063497 Adeno-ass
2	121.8	85.2	4683	14 AF028704	AF028704 Adeno-ass
3	119	83.2	4718	14 AF063497	AF063497 Adeno-ass
4	114.8	80.3	4683	14 AF028704	AF028704 Adeno-ass
5	113.2	79.2	4726	14 AV048704	U48704 Adeno-assoc
6	107.4	75.1	145	6 AR140333	AR140333 Sequence
7	107.4	75.1	145	6 AX286293	AX286293 Sequence
8	107.4	75.1	145	14 AA2LTR2	K01625 Adeno-assoc
9	107.4	75.1	165	6 AR034135	AR034135 Sequence
10	107.4	75.1	165	6 AX106702	AX106702 Sequence
11	107.4	75.1	165	6 I16806	I16806 Sequence 1
12	107.4	75.1	192	6 A46400	A46400 Sequence 3
13	107.4	75.1	198	6 AX418199	AX418199 Sequence
14	107.4	75.1	272	6 AX106701	AX106701 Sequence
15	107.4	75.1	4675	6 AX135805	AX135805 Sequence
16	107.4	75.1	4675	6 AX286292	AX286292 Sequence
17	107.4	75.1	4675	14 AA2CG	J01901 Adeno-assoc
18	107.4	75.1	4679	6 AX282480	AX282480 Sequence
19	107.4	75.1	4679	14 AF043303	AF043303 Adeno-ass
20	107.4	75.1	4680	6 AR028767	AR028767 Sequence
21	107.4	75.1	4680	6 I62303	I62303 Sequence 1
22	107.4	75.1	9258	12 AX205072	AX205072 Sequence
23	107.4	75.1	9258	12 AF383623	AF383623 Cloning v
24	105.8	74.0	191	6 AX398302	AX398302 Sequence
25	105.8	74.0	272	6 AX106699	AX106699 Sequence
26	102.6	71.7	272	6 AX106701	AX106701 Sequence
27	101	70.6	174	6 AX398304	AX398304 Sequence
28	99.4	69.5	145	6 A46401	A46401 Sequence 4
29	99.4	69.5	145	14 AA2LTR1	K01624 Adeno-assoc
30	99.4	69.5	145	14 AA2REPORT	M10681 Adeno-assoc
31	99.4	69.5	165	6 AR034135	AR034135 Sequence
32	99.4	69.5	165	6 AX106702	AX106702 Sequence
33	99.4	69.5	165	6 I16806	I16806 Sequence 1
34	99.4	69.5	2116	14 AA2LEFT	J01902 Adeno-assoc
35	99.4	69.5	4675	6 AX135805	AX135805 Sequence
36	99.4	69.5	4675	6 AX286292	AX286292 Sequence
37	99.4	69.5	4675	14 AA2CG	J01901 Adeno-assoc
38	99.4	69.5	4679	6 AX282480	AX282480 Sequence
39	99.4	69.5	4679	14 AF043303	AF043303 Adeno-ass
40	99.4	69.5	4680	6 AR028767	AR028767 Sequence
41	99.4	69.5	4680	6 I62303	I62303 Sequence 1
42	99.4	69.5	8698	6 AX205072	AX205072 Sequence
43	99.4	69.5	9258	12 AF383623	AF383623 Cloning v
44	98.8	69.1	4722	14 AF028705	AF028705 Adeno-ass
45	98.8	69.1	4726	14 AV048704	U48704 Adeno-assoc

ALIGNMENTS

RESULT 1
LOCUS AF063497 4718 bp DNA linear VRL 27-APR-1999
DEFINITION Adeno-associated virus 1, complete genome.
ACCESSION AF063497
VERSION AF063497.1 GI:4689096
KEYWORDS
SOURCE adeno-associated virus 1.
ORGANISM adeno-associated virus 1
REFERENCE 1 (bases 1 to 4718)
AUTHORS Xiao,W., Chirmule,N., Berta,S.C., McCullough,B., Gao,G. and
TITLE Gene therapy vectors based on adeno-associated virus type 1
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE	99214338
PUBMED	10196295
REFERENCE	2 (bases 1 to 4718)
AUTHORS	Xiao W and Wilson J.M.
TITLE	Direct Submission
JOURNAL	Submitted (05-MAY-1998) HGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA
FEATURES	Location/Qualifiers
Source	1..4718

CDS

```

/product="nonstructural protein"
/protein_id="AAD27758.1"
/db_xref="GI:4689098"
/translation="MPGEFYEIVKPSDLDEHLPGCISDFVSWAEKEMELPSPSDMDI
LNLIEADPLTVAEKLQDFLVQMRVSKAPKLPVQGEKESFYHLHILVETTVGKVS
WYLGFLSQRDKLVQTIYRGIEPLTPMFWATKRNQSGGNKQKVDCEYLPNLLPPL
TQPELQMTNMEETYSACINLAEKRRLVAOHLITVSTOGTOKENKMNPNSDAVPVS
KTSAAFMELVGMVDRGITSSEKQWIOEQDASTISNNAENRSQIKALDNAGKIMAL
TKSADYIVGPAIPADIKTNRYLYRLLENGPEVAGSVFGMQRGKGNKTYLWFE
PATTCTYIAEIAIAHPFVGGCVNTNTEPEPNDQVMVIMWEGKATVSVASAKK
ILGSGKRVDDCKSSAOIDPPTVITSNTNMCAIDGNSSTFEEHQDLOMRKEFEL
TRLRLEDGKATKQEKVEFRMAOHTVEVNAHEFVVRGAGNKPAPADPAKSEKPE
CRSVADPSTSADEGA PVNPDABRYQKCSRHAGMQLMPLPCKTCERNQNPNTCTHGG
RDCCSCFPGVSSQDPVVRKRTYRKLCAIHLHLGRAPETIACACDVLVNDLDCVSEQ

```

CDS

```

translat1on = "MAAGVYIPDMLEUJSLIGSRIMDDKADGAPPKPNAOQKODGGR
LVJGKYKGIAPBNGLDKGEVPDQMLKEDADALHEHDKAYIDQOLKAGNDPVPYRNH4ADFEQBL
RLOEDTSGFCGNLGRAVFOAKKRVLEPLDCLVEEGAKITAGKKRPYBQSGOEFDSSGIGI
KTQGOFPACKRILNFQGTGDSSEVPDPQALPEVPATPAAGPPTMASGGAPADNNEGAP
DGVNAGSANGHMDCSITMLDGRVITTTSTRMALPTVYNHLYKQISASATGASADNHYFGY
SLNPGYVDFPNRHFCHPSPRDMOGLIINNMGRKPKLANKLI NIKVEYETLNDGAVTTLA
NMLSTVQDFNDEYOLPYVLGSAHQGLPPALPANI PYQVGLTLNNGSAOAGRS
FYCLAEYPPSQMLRTGNNTPTSYTPEEYFPHSSYMAISQSLDLNANPLDQYIYLANRTCY
NOSASQONKDLIPGCFRGSFAGMSQPKNMLPGCGTMIHQKSAASATNLDNNTNVTGASL
YNLNGRESI1INPTGAMASHKDEDKFFPMSGGMCHQKSSAKSADNLDNANI1TBEEL
KATNPVATEREGTVAIVNFQSSSTDPTADVHMMGLPMWQODVFLQOQI1WAKI1PHEV
TDGPHHPSPMLMGFELKNPPOILICNTIPVNPALPESAKRKS4FVLYOYSTGVQSVYE
IEMLOKENSKNRNPDEVQYTSNYSKASANDVTDYNDINGLYTERDPR1G1R1YLRPL"

```

	Query Match	Similarity	Score	DB	length
Best Local Match	100.0%	100.0%	143	DB 14	4718
Matches	143	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

RESULT 2			
AF028704/c			
LOCUS	4683 bp	DNA	VRL 12-JAN-1998
DEFINITION	Adeno-associated virus 6, complete genome.	linear	
ACCESSION	AF028704		
VERSION	AF028704.1		
KEYWORDS	GI:2766505		

SOURCE	adeno-associated virus 6.
ORGANISM	adeno-associated virus 6 Virusess; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE	1 (bases 1 to 4683)
AUTHORS	Rutledge,E.A., Halbert,C.L. and Russell,D.W.
TITLE	Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2
JOURNAL	J. Virol. 72 (1), 309-319 (1998)
MEDLINE	98080418
PUBMED	9420229
REFERENCE	2 (bases 1 to 4683)
AUTHORS	Rutledge,E.A. and Russell,D.W.
TITLE	Direct Submision
JOURNAL	Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box 357720, Seattle, WA 98195, USA
FEATURES	Location/Qualifiers
SOURCE	1..4683

CDS

/evidence="not experimental
 /product="nonstructural protein Rep78"
 /protein_id="AAB95449.1"
 /db_xref="org:17266606"
 /translation="MPGFYEVIVIKPSDLDEHLPGISDSFVNVAEKEMLPPDSDDMD
 LNI LEONAPLTAELKLDRLDFVOMWRSEKAPBALPFVQFEKGESEFHLILVETTVGYS
 WYLRSEPLQITRDKLVQTTIYRGIEPTLENFRAVITKRGAGAGKGVDECTIPYLLPK
 TQPIELQAMTMEEYISACINLAKRKLVAHDLVHSYQTEONKEINLPSNDAPVLR
 KTSRYMVELVOMLDRGITSEKQWIOEDQSYISFNASRSGQIKALINDAGTALV
 TKSPADYVGPAPADIKTNRVYIIELENGDPAVYASVFLGAGQKFGKRNITWLE
 PATGTGKNIAEALAHANVPFGYGVAMTNENPPNDUCVKNIMHEEGMTAKVVBESAA
 ILGSKRVDVCKCKSSAIDPDPVITVSNTMCVADIGNSTTFEHQVLDLDRNFKPEL
 TRIRLEFGKTVQKQVEFPFMAQDHTVEAHAEVYVAGANKRPAPDDADKSPKFA
 CSPVADSTISAEAGAPVDPADRYONKCSRAHQIOMLFPCKTCERNQNNFNIETHT
 RDCECFPGVSESPVYRKRTYRKLCAIHHLGRAPLIACSGADLVVAVDLDCCVSEQ"
 2208..4418

CDS

```

/evidence=not experimental
/product="capsid protein VP1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translation="MAADGYLPLWLEDNINSEGIKREWWDLKPGAPKPRANQKQDDRG
LVLPGYKLYPFGNGLDQKRGVFNADAAALAEHDYAPQQLVAGNPGVPIRINHAALFEQ
RLDEDSFGGKILGRAVFAQKARVLPEALGVEEAKTAPEKKRPPVSPSPDSSGIG
KTGQQAQKRLINFGQTDSESVDPDQPLQKPTPAVGPPTMASGGGAPMAINDGEEA
DGVGNASGNHCDSTWLGDRVLTSTRTYMLPTYNHLYKQISSASGANDNHYFEG
STPMGYVEFENRPHCHFSPRDMORLINNNGEPPRKLNPKLFNIOVEVETNDVATYA
NMLTSYQVPSDEYQULPYLGSAGHCCIPPRADYMEIPQYGLTILNNGSQAQVSS
FYCLBYTPSOMLRTGNNFTSTYFEDPHSTASHOSDLRLNMLDLYLTILNTRQ
NQGSGSANKDLIFSRGSPADMSQVOPKRWPGVPCYRQORSKATITDNNNSFWTVAISK
YNNNGESILINPTGMAASHDKDOKFPMGWCVLPKESAGSANTLADNMTIDDEEIL
KANNPVATERFGTVAANLQSSSTDPAQTDVHWAGALPGMWQORDVYLQSPIAKLPH
TQDHFPHSPLMGEGFKLHPSPOLLIKNTLPVAPNAPKSAFTKRSFATVYSTQVSGVE
IMELILKENSKRNNPVOYTSNAKSANDFTVNNGLYTERPRIGIRYLRRL"

```

	Query Match	85.2%;	Score 121.8;	DB 14;	Length 4683;
	Best Local Similarity	98.4%;	Pred. No. 2.3e-18;		
	Matches 123;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	19	TTGGCCACTCCCTCTCTGCGCGCTCGTGCCTGGTGTGGGCGCTTGGGACCAAGTCCCG			78
Db	4683	TTGGCCACTCCCTCTAAGCGCGCTCGTGCCTGGTGTGGGCGCTTGGGACCAAGTCCCG			4624
Qy	79	AGACGCGAGAGCTCTGCTTGC CGGCCCCACCGAGCGAGCGCGCAGAGAGGAGTG			138
Db	4623	AGACGCGAGAGCTCTGCTTGC CGGCCCCACCGAGCGAGCGGAGCGCATTAGAGGAGTG			4564
Qy	139	GGCMA 143			

Db 4563 GGCNA 4559

RESULT 3
LOCUS AF063497/c 4718 bp DNA linear VRL 27-APR-1999
DEFINITION Adeno-associated virus 1, complete genome.
ACCESSION AF063497
VERSION AF063497.1 GI:4689096
KEYWORDS
SOURCE adeno-associated virus 1.
ORGANISM adeno-associated virus 1
REFERENCE 1 (bases 1 to 4718)
AUTHORS Xiao W., Chirmule, N., Betta, S.C., McCullough, B., Gao, G. and Wilson, J.M.
TITLE Gene therapy vectors based on adeno-associated virus type 1
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)
MEDLINE 99214338
PUBMED 10196295
REFERENCE 2 (bases 1 to 4718)
AUTHORS Xiao, W. and Wilson, J.M.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA
location/Qualifiers
1. 4718
/organism="adeno-associated virus 1"
/db_xref="taxon:85106"
/note="AAV1"
335..2206
/codon_start=1
/product="nonstructural protein"
/protein_id="AAD27758.1"
/db_xref="GI:4689098"
/translation="MGGFVIVIKVPSDLDEHLPGISDSFVSWVAEKEMELPPDSMD
LNLIEQAPLVAEKLDQDFVQWRVSKAEALFVQFEKGSYFHLILVETTVGS
MVGRLSQRDLKQVITVIRGIEPTLPNFAVTRKNGAGGKVVDECVIPYLLPK
TOPELOAMTNNMEYISACINLAERKLVNOLTHVGTQBOENKINPNPSDPIVRS
KTSARVLELGMVLDIGITSEKQIODEQASISFNAASRSQIKALDNACKIMAL
PATGKTNIAEALAAHVPFGCVNNTNENPNPCVDKAVIWEBSKMTAKVESAKA
ILIGSKRVDOCKSSAQIDPTVIVTSNNMCVLDGNSITFEHQPLDNRKFEL
TRLRLEHFGVTKQEVKEFRMAQDHTVEAHEFVYRKGGANRPAPDDAKSEPKRA
CPGVADPSTSDAGAPVDPAADRYONKCSHAGMLQMLPCKTCERNQNNNICFTGT
RDSCEFPVSESQPVVRRKTRKLCALHHLGRABEIAACDLVAVDLDVCSQ"
2223..4433
/codon_start=1
/product="capsid protein"
/protein_id="AAD27757.1"
/db_xref="GI:4689097"
/translation="MAADGYLPMLENDLSEGRREMWDLKPGAPKPKANOQKDDRG
LIVPGYKLGPFNGLDKGEFVNAADAALAHDKAYDQAKAGNPYLRYHAHAEFGE
RLQEDTSGFNGLRAVFOAKRVLEPLGLVEBGAKTAPGKRREVSQPPQDSSSGIG
KTGQOPAKKRLNFGQTDSESVDPQPLGEPPTPAVPTMASSGGGAPMANNNEGA
DVGNASGNNHCDSTWLGDRVITTSRTWALPTVNNHLYKOISASAGSANDNHEGY
STPMGVEDPFRFHCHPSRPDMORLINNNNGFRPKRLNPKLPIQVEVTTNDVTTIA
NNLSTVQVPSDSYQLPYVUSAHQCLPPFADVPMITQYGLTLNNSQAVGRSS
FYCLEYFPQMLRTGNNTFSYTEEVFPHSSVAHSQSLRLNNPLIDQVLYLNRO
NOSGSAQNKDLFSGRSPAGMSVQPKMWLPGPCQORVSKTKTDNNNSFTYTGASK
YLNNGRESIINPGTAAASHKDDKFFPMSGVMIFGESAGASNTALDNMTDDEBI
KATNPVATERFGTVAVNFOSSTDPATGVAHAGALPGMWODRDVYLOGPIWAKIPIH
TDGHFHPSPLMGFGGLKHPPIQLIKNTFVPAHPAEFATKPSFTITQYSTGOVSE
IEMELQENSRKMPVEQVYTSNAKSNVDFYDNNGLYTEPRITGRVYLRPL"

BASE COUNT 1121 a 1393 c 1273 g 931 t

ORIGIN

Query Match 83.2%; Score 119; DB 14; Length 4718;
Best Local Similarity 89.5%; Pred. No. 1e-17;
Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TTACCCCTAGTAGTAGTGGCCACTCCCTCTGGCGGCTCGCTCGCTGGTGGGCC 60
|||||

Db 143 TTACCCCTAGTAGTAGTGGCCACTCCCTCTGGCGGCTCGCTCGCTGGTGGGCC 84

QY 61 TGGGACCAAGGTCGGCAGAGCGCAGAGCTGCTGTGCGGGCCACACGAGCGCA 120
|||||

Db 83 GGCAGAGCAGAGCTGCGCTGCGGACCTTGTTCGAGGCCACCGACGAGCGCA 24
|||||

QY 121 GCGCGAGAGAGGAGTGGCGCA 143
|||||

Db 23 GCGCGAGAGAGGAGTGGCGCA 1
|||||

RESULT 4
LOCUS AF028704 4683 bp DNA linear VRL 12-JAN-1998
DEFINITION Adeno-associated virus 6, complete genome.
ACCESSION AF028704
VERSION AF028704.1 GI:2766605
KEYWORDS
SOURCE adeno-associated virus 6.
ORGANISM adeno-associated virus 6
REFERENCE 1 (bases 1 to 4683)
AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2
JOURNAL J. Virol. 72 (1), 309-319 (1998)
MEDLINE 98080418
PUBMED 9420229
REFERENCE 2 (bases 1 to 4683)
AUTHORS Rutledge, E.A. and Russell, D.W.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box 357720, Seattle, WA 98195, USA.
location/Qualifiers
1. 4683
/organism="adeno-associated virus 6"
/db_xref="taxon:68558"
/note="AAV6"
320..2191
/codon_start=1
/evidence=not_experimental
/product="nonstructural protein Rep78"
/protein_id="AAB95449.1"
/db_xref="GI:2766606"
/translation="MGGFVIVIKVPSDLDEHLPGISDSFVSWVAEKEMELPPDSMD
LNLIEQAPLVAEKLDQDFVQWRVSKAEALFVQFEKGSYFHLILVETTVGS
MVGRLSQRDLKQVITVIRGIEPTLPNFAVTRKNGAGGKVVDECVIPYLLPK
TOPELOAMTNNMEYISACINLAERKLVNOLTHVGTQBOENKINPNPSDPIVRS
KTSARVLELGMVLDIGITSEKQIODEQASISFNAASRSQIKALDNACKIMAL
PATGKTNIAEALAAHVPFGCVNNTNENPNPCVDKAVIWEBSKMTAKVESAKA
ILIGSKRVDOCKSSAQIDPTVIVTSNNMCVLDGNSITFEHQPLDNRKFEL
TRLRLEHFGVTKQEVKEFRMAQDHTVEAHEFVYRKGGANRPAPDDAKSEPKRA
CPGVADPSTSDAGAPVDPAADRYONKCSHAGMLQMLPCKTCERNQNNNICFTGT
RDSCEFPVSESQPVVRRKTRKLCALHHLGRABEIAACDLVAVDLDVCSQ"
2208..4418
/codon_start=1
/evidence=not_experimental
/product="capsid protein VP1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translation="MAADGYLPMLENDLSEGRREMWDLKPGAPKPKANOQKDDRG
LIVPGYKLGPFNGLDKGEFVNAADAALAHDKAYDQAKAGNPYLRYHAHAEFGE
RLQEDTSGFNGLRAVFOAKRVLEPLGLVEBGAKTAPGKRREVSQPPQDSSSGIG
KTGQOPAKKRLNFGQTDSESVDPQPLGEPPTPAVPTMASSGGGAPMANNNEGA
DVGNASGNNHCDSTWLGDRVITTSRTWALPTVNNHLYKOISASAGSANDNHEGY
STPMGVEDPFRFHCHPSRPDMORLINNNNGFRPKRLNPKLPIQVEVTTNDVTTIA
NNLSTVQVPSDSYQLPYVUSAHQCLPPFADVPMITQYGLTLNNSQAVGRSS
FYCLEYFPQMLRTGNNTFSYTEEVFPHSSVAHSQSLRLNNPLIDQVLYLNRO
NOSGSAQNKDLFSGRSPAGMSVQPKMWLPGPCQORVSKTKTDNNNSFTYTGASK
YLNNGRESIINPGTAAASHKDDKFFPMSGVMIFGESAGASNTALDNMTDDEBI
KATNPVATERFGTVAVNFOSSTDPATGVAHAGALPGMWODRDVYLOGPIWAKIPIH
TDGHFHPSPLMGFGGLKHPPIQLIKNTFVPAHPAEFATKPSFTITQYSTGOVSE
IEMELQENSRKMPVEQVYTSNAKSNVDFYDNNGLYTEPRITGRVYLRPL"

ISEMLOKENSXRNPEVOYTSNVAKSANVDFVTNNGLYTEPRDGTGRYLRLPL"

BASE COUNT 1114 a 1363 c 1277 g 929 t

ORIGIN

Query Match 80.3%; Score 114.8; DB 14; Length 4683; *
Best Local Similarity 88.0%; Pred. No. 9.3e-17;
Matches 125; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 TACCCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGCTCGGTGGGGGCT 61
Db 4542 TACCCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGCTCGGTGGGGGCT 4601

QY 62 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 121
Db 4602 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 4661

QY 122 CGCGCAGAGAGAGTGGGCA 143
Db 4662 CGCGCAGAGAGAGTGGGCA 4683

RESULT 5 4726 bp DNA linear VRL 15-JUL-1996
AVU48704
LOCUS
DEFINITION Adeno-associated virus 3 nonstructural protein and capsid protein
genes, complete cds, and complete genome.
VERSION U48704
KEYWORDS U48704.1 GI:1408467
SOURCE
ORGANISM Adeno-associated virus 3 strain=3H.
VIRUSES; asDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4726)
AUTHORS Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.
TITLE Nucleotide sequencing and generation of an infectious clone of
JOURNAL adeno-associated virus 3
MEDLINE Virology 221 (1), 208-217 (1996)
PUBMED 96266430
8661429

REFERENCE 2 (bases 1 to 4726)
AUTHORS Muramatsu, S. and Brown, K.E.
TITLE Direct Submision
JOURNAL Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,
NHBLI/NH, 9000 Rockville, MD 20892, USA
FEATURES
source
1. 4726
Location/Qualifiers
/organism="adeno-associated virus 3"
/strain="3H"
/db_xref="taxon:46350"
318. 2192
/codon_start=1
/product="nonstructural protein"
/protein_id="AAC55048.1"
/db_xref="GI:1408468"
/translation="MPEGYEIVLVKPSDLDERLKIISNSFVNWAKEKWDVPPDSND
PNLEOAPLVAEKLOREFLVEMRVSAPALFVQFEKETYFHLVLIETIGVS
MVRKYVSOIKELVTRIRGVBPOLPFWATKRGNGAGGKAVDDCVIPLYLK
TQPELOAMATMDQVLYACINLAEKRLVAOHLTVSGTOGKONKNNPNSDAVRS
KTSARMEVLVMDRGTTSKQWIOEQAISIPNAASNSRSITKALDASIKIMSL
TKTAPDVIVGSPEDITKRIYQILENGDPOYASVFLGMAOKFRGKNTLWFG
PATTGKTNIAEIAHAYVFCVVMVTENFPNDQVDMVIMWEEGKMTAVVSAIA
ILGSKYRVDQKSSAIEPTPIVTSNTNCVAVIDNSTTFHQPLDQRMPEFL
TRLDHDEGKVTQKQVDFPMASDHVDVHEFVVRGAKKRPASNDADVSPKE
QTSIAOPTSDAEPADYADRYOKKCSHVGMNMLPCKTCERMMNOISNCPFHGR
DCEGCFRMSSESQPVSVKKTYYKTLCTIHHILGAPFIACADLANVDDCVSBO
2209. 4419
/codon_start=1
/product="capsid protein"
/protein_id="AAC55049.1"
/db_xref="GI:1408469"
/translation="MAAGYLPDMLEDNLSGIRBWMALKGVDPQRYANOQHQRNG
LVLPQYKTLGPNGILDKEGPVNEADAAALEHDKAYDOLKAGDNPYLYKNHADAEPOR

RLOEDTSFGGILGRAVFOAKRIILEPLGLVEBAKTAPEGKAGVDSPOEPDSSGCV
KSGKOPARKRLINFGOTDSSEVPDPOLEGPAPAPISLSGNTMAGSGAGMANNEGA
DGVGNSGNNHNDOSWILGDRVITTSRTALPTYNNHLYOISOGSAGNDNHYFGS
TPWGYDPRFRFHFRFBPRMQRILNNMGRPKLSFKLFNIOVRYTONDCTTTAN
NLSTVOVFTDSYOLPYVIGSAHQGLPEFPADVPVPOYGLTLNNGQAYGRSSE
YCLEYFSPQMLRRTGNNEPQFSTFEDVFPFSSVAHOSLDLNPILIDQYLYNRTQG
TTGTTNOSHLFSQAGPOSMSLOARNMLPGPCYRQRLSKTNDNNNSNPMPTASK
YHLNGRSLVNPGRPAMASHKDEKEFPFMGNLI FGKETTASNAELDNVMTDEE
RTNPPATVEGYVANNLOSNTAPPTGTNHOAGALPGMWODRDVYLOGPIWAKI
TDGHPHSPLEMGFRGLKHPBPQIMIKNTVPYRANPTTSPAKRASFITQVTSQVSE
ISEMLOKENSXRNPEIQTYSNYSNVNDFVTNNGVSEPRPIGRYLRLPL"

BASE COUNT 1200 a 1295 c 1251 g 980 t

ORIGIN

Query Match 79.2%; Score 113.2; DB 14; Length 4726;
Best Local Similarity 87.3%; Pred. No. 2.1e-16;
Matches 124; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 TACCCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGGTGGGGGCT 61
Db 4585 TACCCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGGTGGGGGCT 4644

QY 62 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 121
Db 4645 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 4704

QY 122 CGCGCAGAGAGAGTGGGCA 143
Db 4705 TGCGCATGAGAGAGTGGGCA 4726

RESULT 6 145 bp DNA linear PAT 16-JUN-2001
AR14033/c
LOCUS
DEFINITION Sequence 1 from patent US 6207457.
ACCESSION AR14033
VERSION AR14033.1 GI:14482829
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 145)
AUTHORS Matsushita, G. and Surosky, R.T.
TITLE Targeted nucleotide sequence delivery and integration system
JOURNAL Patent: US 6207457-A 1 27-MAR-2001;
FEATURES
source
1. 145
Location/Qualifiers
/organism="unknown"
BASE COUNT 20 a 52 c 49 g 24 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 145;
Best Local Similarity 85.1%; Pred. No. 9.3e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGGTGGGGGCT 62
Db 141 ACCCTAGTATGAGTGGCCCTCCCTCGCGCGGCTCGCTCGGTGGGGGCT 82

QY 63 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 122
Db 81 GCGACCAAGAGTCCGACAGAGAGCTCTGCTCTCGCGCGGCTCGCTCGGTGGGGGCT 22

QY 123 GCGCAGAGAGAGTGGGCA 143
Db 21 GCGCAGAGAGAGTGGGCA 1

RESULT 7 145 bp DNA linear PAT 21-NOV-2001
AX286293
LOCUS
DEFINITION Sequence 2 from Patent W00180840.
ACCESSION AX286293

VERSION AX286293.1 GI:17048541
KEYWORDS adeno-associated virus 2.
ORGANISM adeno-associated virus 2.
REFERENCE 1
AUTHORS Raj, K. and Beard, P.M.
TITLE Cytotoxic agents
JOURNAL Patent: WO 0180840-A 2 01-NOV-2001;
BTG INTERNATIONAL LIMITED (GB)
FEATURES
source
1. .145
/organism="adeno-associated virus 2"
/db_xref="taxon:10804"
misc_structure 1. .145
/note="ITR"
misc_feature 72
/note="unpaired base"
misc_feature 94
/note="unpaired base"
BASE COUNT 24 a 49 c 52 g 20 t
ORIGIN
Query Match 75.1%; Score 107.4; DB 6; Length 145;
Best Local Similarity 85.1%; Pred. No. 9.3e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 5 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 64
QY 63 CGGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCCACCGAGCGAGCG 122
DB 65 GCGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCTCACTGAGCGAGCG 124
QY 123 GCGCAGAGAGGAGTGGCCAA 143
DB 125 GCGCAGAGAGGAGTGGCCAA 145
RESULT 8
AA2LTR2
LOCUS 145 bp DNA linear VRL 27-APR-1993
DEFINITION Adeno-associated virus 2 right terminal sequence.
ACCESSION K01625
VERSION K01625.1 GI:209624
KEYWORDS replication; terminal repeat.
SEGMENT 2 of 2
SOURCE Adeno-associated virus 2H DNA (clone pSM620 [2]), from KB or HeLa cells.
ORGANISM adeno-associated virus 2H
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 145)
AUTHORS Lusby, E., Fife, K.H. and Berns, K.I.
TITLE Nucleotide sequence of the inverted terminal repetition in
JOURNAL adeno-associated virus DNA
MEDLINE J. Virol. 34 (2), 402-409 (1980)
PUBMED 80185149
REFERENCE 2 (bases 1 to 145)
AUTHORS Lefebvre, R.B., Riva, S. and Berns, K.I.
TITLE Conformation takes precedence over sequence in adeno-associated
JOURNAL virus DNA replication
MEDLINE Mol. Cell. Biol. 4 (7), 1416-1419 (1984)
PUBMED 85061247
COMMENT
The focus of both papers is the method of replication of the virus.
[1] notes that the initial it is present only 30% of the time; it
is shortened to t in 50% of the population and missing altogether
in 15% of the population. There is further sequence heterogeneity
which can be explained by assuming that the terminal 125 bases,
which form an imperfect palindrome, are replaced by their inverted
complement during replication. [2] found that deletion of the 9

terminal bases on the right and the 113 terminal bases on the left
of AAV 2 genome did not stop DNA replication. Further deletion of
an 11-base symmetrical sequence (bases 89 to 99) in the right
terminal repetition inhibits DNA replication. Substitution of
either an 8-base (cagatctg) or 12-base (cgcgagccgcg) symmetrical
sequence unrelated to the original 11-base sequence restores DNA
replication. All of this can be explained by assuming that the 125
base palindrome mentioned above form a t-shaped secondary structure
which provides a primer for DNA polymerase during replication.
FEATURES
source
1. .145
/organism="adeno-associated virus 2H"
/db_xref="taxon:10805"
BASE COUNT 24 a 49 c 52 g 20 t
ORIGIN
Query Match 75.1%; Score 107.4; DB 14; Length 145;
Best Local Similarity 85.1%; Pred. No. 9.3e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 5 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 64
QY 63 CGGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCCACCGAGCGAGCG 122
DB 65 GCGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCTCACTGAGCGAGCG 124
QY 123 GCGCAGAGAGGAGTGGCCAA 143
DB 125 GCGCAGAGAGGAGTGGCCAA 145
RESULT 9
AR034135
LOCUS 165 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5869305.
ACCESSION AR034135
VERSION AR034135.1 GI:5949740
KEYWORDS
ORGANISM Unknown.
SOURCE Unknown.
REFERENCE 1 (bases 1 to 165)
AUTHORS Samulski, R. Jude, and Xiao, X.
TITLE Recombinant viral vector system
JOURNAL Patent: US 5869305-A 1 09-FEB-1999;
FEATURES Location/Qualifiers
1. .165
/organism="unknown"
BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN
Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 9.1e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 3 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 62
DB 5 ACCCTTAGTATGATGAGTGGCCCACTCCCTCTCTGCGGCGCTGCTGCTGCGTGGGCGCTG 64
QY 63 CGGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCCACCGAGCGAGCG 122
DB 65 GCGACCAAGAGTCCGACAGCGGAGAGTCTGCTGCTGCGGCGGCGCTCACTGAGCGAGCG 124
QY 123 GCGCAGAGAGGAGTGGCCAA 143
DB 125 GCGCAGAGAGGAGTGGCCAA 145
RESULT 10
AX106702
LOCUS 165 bp DNA linear PAT 30-APR-2001

DEFINITION Sequence 7 from Patent WO0125465.
ACCESSION AX106702
KEYWORDS AX106702.1 GI:13922363
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 165)
AUTHORS Engelhardt, J. F., Dongsheng, D. and Ziyang, Y.
TITLE Adeno-associated viruses and uses thereof
JOURNAL Patent: WO 0125465-A 7 12-APR-2001;
University of Iowa Research Foundation (US) ; Engelhardt, John F.
(US) ; Dongsheng, Duan (US) ; Ziyang, Yan (US)
LOCATION/Qualifiers
1..165
/organism="unidentified"
/db_xref="taxon:32644"
/note="SEQ ID NO:1 of U.S. Patent No. 5,478,745"

BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 9.1e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTTGCCACCTCCCTCTGCGCGCTGCTGCTCGGTGGGGGCTG 62
Db 5 ACCCTAGTATGATGAGTTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGCTGAGGCGG 64
Qy 63 GCGACCAAGGTCCGACAGCGGACAGAGTCTGCTCTGCGCGGCCCGACCGAGCGAGC 122
Db 65 GCGACCAAGGTCCGCGCGCGCGGCTTTGGCCCGGCGGCTCAGTAGAGCGAGCGAGC 124
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGGCAA 145

RESULT 11
LOCUS I16806 165 bp DNA linear PAT 03-APR-1996
DEFINITION Sequence 1 from patent US 5478745.
ACCESSION I16806
VERSION I16806.1 GI:1251714
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 165)
AUTHORS Samulski, R.J. and Xiao, X.
TITLE Recombinant viral vector system
JOURNAL Patent: US 5478745-A 1 26-DEC-1995;
FEATRES Location/Qualifiers
1..165
/organism="unknown"

BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 9.1e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTTGCCACCTCCCTCTGCGCGCTGCTGCTCGGTGGGGGCTG 62
Db 5 ACCCTAGTATGATGAGTTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGAGGCGG 64
Qy 63 GCGACCAAGGTCCGACAGCGGACAGAGTCTGCTCTGCGCGGCCCGACCGAGCGAGC 122
Db 65 GCGACCAAGGTCCGCGCGCGCGGCTTTGGCCCGGCGGCTCAGTAGAGCGAGCGAGC 124
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGGCAA 145

RESULT 12
LOCUS A46400 192 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 3 from Patent WO9523867.
ACCESSION A46400
VERSION A46400.1 GI:2300601
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 192)
AUTHORS Deneffe, P., Lacta, M., Perricaudet, M. and Vigne, E.
TITLE INTEGRATIVE RECOMBINANT ADENOVIRUSES, PREPARATION THEREOF AND
JOURNAL THERAPEUTICAL USES THEREOF
Patent: WO 9523867-A 3 08-SEP-1995;
RHONE-POULENC RORER SA (FR)
Other publication AU 1852695 950918
Other publication FR 2716893 950908.
LOCATION/Qualifiers
1..192
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 39 a 58 c 63 g 32 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 192;
Best Local Similarity 85.1%; Pred. No. 8.6e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTTGCCACCTCCCTCTGCGCGCTGCTGCTCGGTGGGGGCTG 62
Db 52 ACCCTAGTATGATGAGTTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGAGGCGG 111
Qy 63 GCGACCAAGGTCCGACAGCGGACAGAGTCTGCTCTGCGCGGCCCGACCGAGCGAGC 122
Db 112 GCGACCAAGGTCCGCGCGCGCGGCTTTGGCCCGGCGGCTCAGTAGAGCGAGCGAGC 171
Qy 123 GCGCAGAGAGGAGTGGGCAA 143
Db 172 GCGCAGAGAGGAGTGGGCAA 192

RESULT 13
LOCUS AX418199/c 198 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 1 from Patent WO0192551.
ACCESSION AX418199
VERSION AX418199.1 GI:21523203
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Samulski, R.J. and McCarty, D.M.
TITLE Duplexed parvovirus vectors
JOURNAL Patent: WO 0192551-A 1 06-DEC-2001;
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL (US)
FEATRES Location/Qualifiers
1..198
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Inverted terminal repeat from the AAV-2 vector
plasmid pSub 201"

BASE COUNT 34 a 65 c 58 g 41 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 198;
Best Local Similarity 85.1%; Pred. No. 8.7e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 ACCCTAGTATGATGAGTTGCCACCTCCCTCTGCGCGCTGCTGCTCGGTGGGGGCTG 62

Db 141 ACCCTAGTATGAGTGGCCACTCCCTCTGCGGCTCGCTCACTAGAGCGG 82
QY 63 CGGACCAAGTCCGCGAGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 122
Db 81 GCGACCAAGTCCGCGAGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 22
QY 123 GCGCAGAGAGAGTGGGCAA 143
Db 21 GCGCAGAGAGAGTGGGCAA 1

RESULT 14

LOCUS AX106701 272 bp DNA linear PAT 01-MAY-2001
DEFINITION Sequence 6 from Patent W00125465.
ACCESSION AX106701
VERSION AX106701.1 GI:13922362
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 272)
AUTHORS Engelhardt, J.F., Dongsheng, D. and Ziyang, Y.
TITLE Adeno-associated viruses and uses thereof
JOURNAL Patent: W0 0125465-A 6 12-Apr-2001;
University of Iowa Research Foundation (US); Engelhardt, John F.
(US); Dongsheng, Duan (US); Ziyang, Yan (US)
Location/Qualifiers
1..272
/organism="unidentified"
/db_xref="taxon:32644"
/note="NAV circular intermediate, clone p1202"

FEATURES
source

BASE COUNT 58 a 78 c 78 g 58 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 272;
Best Local Similarity 85.1%; Pred. No. 8.2e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTGCGGCTCGCTCGAGGCGCTG 62
Db 53 ACCCTAGTATGAGTGGCCACTCCCTCTGCGGCTCGCTCGAGGCGCTG 112
QY 63 CGGACCAAGTCCGCGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 122
Db 113 GCGACCAAGTCCGCGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 172
QY 123 GCGCAGAGAGAGTGGGCAA 143
Db 173 GCGCAGAGAGAGTGGGCAA 193

RESULT 15

LOCUS AX135805 4675 bp DNA linear PAT 29-MAY-2001
DEFINITION Sequence 5 from Patent W00132711.
ACCESSION AX135805
VERSION AX135805.1 GI:14272040
KEYWORDS
SOURCE adeno-associated virus 2.
ORGANISM adeno-associated virus 2
adeno-associated virus 2
viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.

REFERENCE 1 (bases 1 to 4675)
AUTHORS Hermoat, P.U.
TITLE Repo-associated virus aav rep78 major regulatory protein, mutants thereof and uses thereof
JOURNAL Patent: W0 0132711-A 5 10-MAY-2001;
THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS (US)
Location/Qualifiers
1..4675
/organism="adeno-associated virus 2"
/db_xref="taxon:10804"

FEATURES
source

CDS

321..2186

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC39573.1"

/db_xref="GI:14272041"

/translation="MPEFYELVLYKPSDLGLHGLISDSFVNVVAKEWELPPDSMD
LNLISQAPLVAEKLDQRLDTEWRVRSKAPALFVQPEKSSVFPNHYLVETTVGS
MVLGRFLSQRLEKLDQRLDTEWRVRSKAPALFVQPEKSSVFPNHYLVETTVGS
TOPELOMANTNMNEQYLSACLNTLTERKLYAHLTHVSTQPEKKNRNSDAPVRS
KTSARYMELVGLVNDKGTISSEKOWIOEDASTYSFNAANSRSQITAAIDNAGKIMSL
TKTAPYLVGQOPEVEDISSNRYKILDELNGYDQYASVFLGWATKFGKRNITWLG
PATGKTNIAEALAHVTPFGCVNMTNENFPENDCVDAKVIWMEEGKMTAKVESA
ILGSKVRDCKSSAOIDPEPVIYTSNTNMCAYIDGASTTEHOOPLODRMEKEL
TRRLDHDGKVTQKQEVKDFPRRAKDHVEVEHFKKGAKKRPA.PSADLSEPRV
RESVAPSTSDAASINVDRIYONKSRHYGNALMFPKQCEPNNQNSNICTHQK
DCLCEPVESSQPSVVKAYQKLCYIHHIMKVPDAGTACDLVNDLDCIFEQ"

BASE COUNT 1198 a 1262 c 1251 g 964 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 4675;
Best Local Similarity 85.1%; Pred. No. 4.5e-15;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTGCGGCTCGCTCGAGGCGCTG 62
Db 4535 ACCCTAGTATGAGTGGCCACTCCCTCTGCGGCTCGCTCGAGGCGCTG 4594
QY 63 CGGACCAAGTCCGCGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 122
Db 4595 GCGACCAAGTCCGCGAGAGAGTCTGCTCTCCGCGCCACCGAGCGAGC 4654
QY 123 GCGCAGAGAGAGTGGGCAA 143
Db 4655 GCGCAGAGAGAGTGGGCAA 4675

Search completed: July 20, 2003, 13:19:54
Job time : 378.075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 212.5 Seconds

(without alignments)
10898.630 Million cell updates/sec

Title: US-09-807-802a-1_COPY_4576_4718

Perfect score: 143

Sequence: 1 ttaccctaagtatgagagtc.....cgacagagagagtgagca 143

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST: *

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hcc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hcc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.8	26.4	763	12	BG490707 602520211
C 2	37.4	26.2	644	10	BB632576 BB632576
C 3	36.2	25.3	604	13	BI531460 BI531460
C 4	36.2	25.3	932	17	AL066742 Drosophila
C 5	35.6	24.9	578	10	BB651469 BB651469
C 6	35.6	24.9	793	13	BI905802 BI905802

C 7	35.6	24.9	1115	14	BM906771
C 8	35.4	24.8	1123	17	AG080476
C 9	35.2	24.6	2024	17	AG074763
C 10	35	24.5	844	17	CNS0052P
C 11	34.6	24.2	860	12	BF578052
C 12	34.4	24.1	890	17	AG127787
C 13	34.4	24.1	1136	13	BM416106
C 14	34.2	23.9	914	13	BI762701
C 15	34.2	23.9	1009	10	BB61010W
C 16	34	23.8	623	10	BB613613
C 17	34	23.8	639	10	BB651977
C 18	34	23.8	643	10	BB652152
C 19	34	23.8	657	10	BB651054
C 20	34	23.8	661	10	BB621081
C 21	34	23.8	795	17	AO331299
C 22	33.8	23.6	881	17	CNS031TR
C 23	33.8	23.6	932	17	CNS0072Q
C 24	33.8	23.6	1372	13	BI913462
C 25	33.6	23.5	1020	17	CNS004YU
C 26	33.6	23.5	1136	13	BM416106
C 27	33.4	23.4	468	9	AA043429
C 28	33.4	23.4	655	13	BI907522
C 29	33.4	23.4	683	13	BI768775
C 30	33.4	23.4	891	13	BI906775
C 31	33.4	23.4	1047	14	BM924938
C 32	33.4	23.4	1284	9	BM547577
C 33	33.2	23.2	283	3	AA874212
C 34	33.2	23.2	879	17	AG141413
C 35	33	23.1	901	12	BF316571
C 36	33	23.1	923	14	BO718248
C 37	32.8	22.9	445	9	AL514067
C 38	32.8	22.9	925	17	CNS0091P
C 39	32.6	22.8	298	17	CNS04MX
C 40	32.6	22.8	727	13	BI454382
C 41	32.6	22.8	877	17	AG127610
C 42	32.6	22.7	984	12	BG809678
C 43	32.4	22.7	337	10	BE110673
C 44	32.4	22.7	787	17	AG133411
C 45	32.4	22.7	822	17	AG134283

ALIGNMENTS

RESULT 1
BG490707/c 763 bp mRNA linear EST 27-MAR-2001
LOCUS 602520211F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4638699 5',
DEFINITION mRNA sequence.
ACCESSION BG490707
VERSION BG490707.1 GI:13452217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 763)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straubeberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM401 row: e column: 04
High quality sequence stop: 182.
Location/Qualifiers
1..763

FEATURES
source

BASE COUNT 188 a 241 c 310 g 24 t

ORIGIN

Query Match 26.4%; Score 37.8; DB 12; Length 763;
Best Local Similarity 58.4%; Pred. No. 6.3;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Dy 26 CTCCTCCTGCGGCCTGCTGCTGCGGGAGCCGTGGGACCAAAAGGCCGACAGCC 85
Dd 261 CGCTGCGCCTCCGCGGCGCTGCTGCTGCTTCCCTTCGCCGCTTGCGCTTCCGCC 202

Dy 86 AGAGCTCTGCTCTGCGGCCCCCACCGACGACGAGCGCGCAGAGAGGATG 138
Dd 201 AGCGCTACGCGCTTCCGCGGCGCTGACGCGCGAGGAGCGGCGCTGCTGCGCGCTG 149

RESULT 2
LOCUS BB632576 644 bp mRNA linear EST 26-OCT-2001
DEFINITION BB632576 RIKEN full-length enriched, adult male hypothalamus Mus musculus CDNA clone A230067E15 5', mRNA sequence.
ACCESSION BB632576
VERSION BB632576.1 GI:16469164
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 644)

REFERENCE Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Komno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ono,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakhi,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

AUTHORS MURAMATSU,M. and HAYASHIZAKI,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shiohito-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length CDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagui,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wachihi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format Sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

	FEATURES									
	SOURCE	<p>Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Caminici,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11(2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamakake,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details.</p>								
	Location/Qualifiers	<pre> 1..644 /organism="Mus musculus" /db_xref="taxon:10990" /cclone_id="A230067R15" /cclone_id="RIKEN full-length enriched, adult male hypothalamus" /sex="male" /tissue_type="hypothalamus" /dey_stage="adult" /lab_host="DH10B" /note="Site.1: Salt; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'. GAGGAGAAGAGATCCAAAGCTCTTTTTTTTTTTTTN 3'). cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGAGAAGATTCTCGGTAAATTAATTAATTAATTAATTA modified pluescript KS(+) after bulk excision from Lambda FLC I." </pre>								
	BASE COUNT	97	a	245	c	194	g	108	t	
	ORIGIN									
	Query Match	26.2%; Score 37.4; DB 10; Length 644;								
	Best Local Similarity	64.4%; Pred. No. 7.5;								
	Matches	56;	Conservative	0;	Mismatches	31;	Indels	0;	Gaps	0;
OY		29	CCTCTTCGCCGCCTCGCTCGCTCGATGGAGGCGCTCGACCAAGATCCGACAGCGCAGA	88						
Db		535	CCTTCGCGCGCCCTGGCGCAGGTCTGTGGCCGCGCGGACAGAACACGACACGCCCATG	594						
OY		89	GCTCTGCTCTGCGCGCCCCCACCGACGC	115						
Db		595	CCTCTACTTGACGTGCCCCCGCGGTG	621						
	RESULT 3									
BIS31460/c	LOCUS	BIS31460 604 bp mRNA linear EST 29-AUG-2001								
DEFINITION		1024113H09.y1 C. reinhardtii CC-1690, normalized, Lambda zap II								
ACCESSION		Chlamydomonas reinhardtii cDNA, mRNA sequence.								
VERSION		BIS31460								
KEYWORDS		BIS31460.1 GI:15372034								
SOURCE		EST.								
ORGANISM		Chlamydomonas reinhardtii.								
		Chlamydomonas reinhardtii								
		Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;								
		Chlamydomonadaceae; Chlamydomonas.								
REFERENCE		1 (bases 1 to 604)								
AUTHORS		Grosman,A., Chang,C.-W., Davies,J.J., Harris,E., Hauser,C., Lefebvre,								
TITLE		P., McDermott,J.P., Strager,J., Stifflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model,								

Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1024b
Unpublished (2001)
Contact: Charles Hauser
Duke Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chausere@duke.edu.

FEATURES
source
Location/Qualifiers

1. .604
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"

/note="Vector: Bluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
Bluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with Exsist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldi et al (1996) Genome Research 6: 791-806."

BASE COUNT 141 a 181 c 213 g 69 t

ORIGIN

Query Match 25.3%; Score 36.2; DB 13; Length 604;
Best Local Similarity 57.5%; Pred. No. 15;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

16 GAGTGGCCCACTCCCTCTGTCGCGCTCGCTGCGGCGCTGCGGACCAAGATC 75
Db 392 GCGTCTCTTCTTCTCTCTGCGCCCATCTCCGTGGATGCGGCTGCGCGCTC 333

76 CGCAGACGCGAGAGCTGCTGCTGCGCGCCCGACGAGCGAGCGCGCAG 128
Db 332 GCGGCGCGCGAGCGCGCGCGCTGCGCTGCGCGCTGCGGAGCTG 280

RESULT 4 932 bp DNA linear GSS 03-JUN-1999
CNS00720
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14B09 of RPCR-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 932)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 131 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Ooeagawa and
Aaron Mammoeer in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCR-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
Location/Qualifiers

1. .932
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACR14B09"
/clone_lib="RPCR-98"
/note="end : T7"

BASE COUNT 155 a 202 c 241 g 91 t 243 others

ORIGIN

Query Match 25.3%; Score 36.2; DB 17; Length 932;
Best Local Similarity 36.3%; Pred. No. 17;
Matches 41; Conservative 32; Mismatches 40; Indels 0; Gaps 0;

22 CCCACTCCCTCTCTGCGCGCTCGCTGCGTGGGCGCTGCGACCAAGTCCGAGA 81
Db 725 SSCGCGCCCGCGCCCGCCCGCGSSGCGGCGCGCGCGCGCGCGCGCG 784

82 CGCAGAGCTCTCTCTGCGCGCGCCCGACGAGCGAGCGCGCGAGAGGG 134
Db 785 CGGCG 837

RESULT 5 578 bp mRNA linear EST 26-OCT-2001
BB651469
LOCUS
DEFINITION
BB651469 RIKEN full-length enriched, 0 day neonate cerebellum Mus-
musculus cDNA clone C230083H05 5', mRNA sequence.
BB651469
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 578)

REFERENCE
AUTHORS
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Komuro, H., Kouda,
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsukuba, Ibaraki, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayashizaki, N., Sugahara, Y., Shibata, K., Itoh,
M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagti, K., Fujisaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer. Genome Res.

10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamakake, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source

1. 578
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230083H05"
/clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTATTAATTAATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pbluescript KS(+) after bulk excision from lambda PUC I."

BASE COUNT 99 a 217 c 147 g 112 t 3 others
ORIGINQuery Match 24.9%; Score 35.6; DB 10; Length 578;
Best Local Similarity 64.6%; Pred. No. 21;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 29 CCTCTCTGCGGCTCGCTCGCTGCTGCGGCGCTCGGACCAAGATCCGACGAGCGAGA 88
|||
Db 22 CCTCTCGCGCGCTCGGCGAGGTGGGGCGCGCGGCGGACAGACGACGACGCGCATG 81
|||
Qy 89 GCTTGTCTGCGGCGGCGCCAC 110
|||
Db 82 CCTCTACTTTGACGCTGCCACCC 103
|||

RESULT 6
BI905802/c 793 bp mRNA linear EST 16-OCT-2001
LOCUS 603062735F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212016 5',
DEFINITION mRNA sequence.
ACCESSION BI905802
VERSION BI905802.1 GI:16168381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;
1 (bases 1 to 793)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
Plate: L1AM1532 row: e column: 09
High quality sequence stop: 780.
Location/Qualifiers
1. 793

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5212016"
/clone_lib="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

BASE COUNT 121 a 246 c 254 g 172 t
ORIGINQuery Match 24.9%; Score 35.6; DB 13; Length 793;
Best Local Similarity 55.7%; Pred. No. 24;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 7 CTAGTATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGTGGGCGCTTGGCA 66
|||
Db 641 CGAGTGTGGGAGCTTAATCCATTGCTCTAGCTTAATCTGGGTAGTCACTGA 582
|||
Qy 67 CCAAGTCCGCGACGCGAGCTCTGCTGCGGCGCGCCACCGACGCGAGCGCGC 126
|||
Db 581 GCACGGCGCCACGCGACGAGCTGTGCACCGCGCGCTCAGTTCAGGCGAGCTCAC 522
|||
Qy 127 AG 128
|||
Db 521 AG 520
|||

RESULT 7
BM906771/c 1115 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT_6621680 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5724802
DEFINITION 5', mRNA sequence.
ACCESSION BM906771
VERSION BM906771.1 GI:19357150
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;
1 (bases 1 to 1115)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>

Plate: LHAM12714 row: k column: 11
High quality sequence start: 9
High quality sequence stop: 691.
Location/Qualifiers

FEATURES

source

1. 1115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5724802"
/clone_1lb="NIH MGC_125"
/lab_host="DH10B"
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1.3-5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."

BASE COUNT 191 a 346 c 325 g 253 t

ORIGIN

Query Match 24.9%; Score 35.6; DB 14; Length 1115;
Best Local Similarity 55.7%; Pred. No. 26;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 7 CTAGTATGAGTGGCCACCTCCCTCTGCGCGCTGCTGCGTGGGGGCTGGCGA 66
DB 636 CGAGTGTGGAGCTTAAATCCATTGCTCTGCTTAAATTCGGGTAGTACGCTGA 577
QY 67 CCAAGGTCCGACGACGCGAGCTGCTGCGCGCCCGACCGAGCGAGCGCGC 126
DB 576 GCACGGCGCCACGCGACGAGCTGTGACCGCGCGCTCGAGGTCCAGGCGAGCTCAC 517
QY 127 AG 128
DB 516 AG 515

RESULT 8
LOCUS AG080476 1123 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-076K10.R, genomic survey sequence.
ACCESSION AG080476
VERSION AG080476.1 GI:16632278
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
BAC library clone:PTB-076K10.R.
ORGANISM Pan troglodytes

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of library PTB

UNPUBLISHED 2 (bases 1 to 1123)

REFERENCE 2 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission

REFERENCE Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

REFERENCE 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

REFERENCE Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145
R.Site 1 : SacI

R.Site 2 : SacI.
Location/Qualifiers
1. 1123
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-076K10.R"
/sex="male"
/cell_type="lymphoblast"
/clone_1lb="PTB Chimpanzee Male BAC library"

BASE COUNT 38 a 403 c 570 g 29 t 83 others

ORIGIN

Query Match 24.8%; Score 35.4; DB 17; Length 1123;
Best Local Similarity 57.1%; Pred. No. 30;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 34 CTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 93
DB 487 CG 546
QY 94 GCTCTGCG 138
DB 547 GCGCCNCG 591

RESULT 9
LOCUS AG074763 2024 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-066P10.F, genomic survey sequence.
ACCESSION AG074763
VERSION AG074763.1 GI:16626565
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
BAC library clone:PTB-066P10.F.
ORGANISM Pan troglodytes

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of library PTB

UNPUBLISHED 2 (bases 1 to 2024)

REFERENCE 2 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission

REFERENCE Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

REFERENCE 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

REFERENCE Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1. 2024

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-066P10.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1lb="PTB Chimpanzee Male BAC library"

BASE COUNT 542 a 537 c 704 g 83 t 158 others

ORIGIN

Query Match 24.6%; Score 35.2; DB 17; Length 2024;
Best Local Similarity 59.2%; Pred. No. 40;

Matches	58:	Conservative	0:	Mismatches	40:	Indels	0:	Gaps	0:
Qy	36	GGCGGCTGCTTCGCTCGGTGGGCGCTCGCGACCAAGTCCGACCGCAGAGCTTCG	95						
Db	1047	GGCGNCACGCGCGCGACCGCGACCGGGGACAGGCCCGGAGAAGGCCCTCGNAGAGACGCGC	1106						
Qy	96	TCTGCGGACCCACCGACGACGACGCGCGACAGAGG	133						
Db	1107	GGCGCGCGCGCGACGACGCGCGACGCGCGACGAGG	1144						
RESULT 10									
CNS0052P	844 bp	DNA	linear	GSS	03-JUN-1999				
LOCUS									
CNS0052P									
DEFINITION	Drosophila melanogaster genome survey sequence TERT end of BAC #								
LOCUS	BACR1P16 of RPCT-98 library from Drosophila melanogaster (fruit								
FLY1), genomic survey sequence.									
AL056652									
AL056652.1	GI:4932342								
GSS.									
Drosophila melanogaster.									
Drosophila melanogaster									
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;									
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;									
Ephydroidea; Drosophilidae; Drosophila.									
1 (bases 1 to 844)									
Genoscope.									
Direct Submission									
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage									
BP 191 91006 EVRI cedex - FRANCE (E-mail : segret@genoscope.cns.fr									
- Web : www.genoscope.cns.fr)									
Determination of this BAC-end and sequence was carried out as part of a									
collaboration with the Berkeley Drosophila Genome Project (BDGP).									
The BDGP is constructing a physical map of the Drosophila									
melanogaster genome using these BACs. For further information									
please see http://www.fruitfly.org The BDGP Drosophila									
melanogaster BAC library was prepared by Kazuhiro Osoegawa and									
Aaron Mammeter in Pieter de Jong's laboratory in the Department of									
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,									
NY. The library is named RPCT-98 and was constructed by partial									
ECORI digestion of Drosophila DNA provided by the BDGP from the									
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's									
PL and EST libraries. A more detailed description of the library									
and how to order individual BAC clones, the entire library, or									
filters for hybridization from the BACPAC Resource Center can be									
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .									
Location/Qualifiers									
1. 844									
/organism="Drosophila melanogaster"									
/db_xref="taxon:7227"									
/clone="BACR1P16"									
/clone_11b="RPCT-98"									
/note="end : TERT3"									
BASE COUNT	261 a	112 c	92 g	35 t	344 others				
ORIGIN									
Query Match	24.5%;	Score 35;	DB 17;	Length 844;					
Best Local Similarity	21.5%;	Pred. No. 35;							
Matches	26;	Conservative 61;	Mismatches 34;	Indels 0;	Gaps 0;				
Db	490	SMCGCMCSVSVCAVCGSSGSRVAVCGSVGGRSSRGHAGSSSGRGSSSVSSGV	549						
Qy	81	ACGCGAGACTCTGCTTCGCGCGGCCACCGACGACGACGACGACGAGTGGG	140						
Db	550	SSSSSVGMCACSSASVSVCBSVSASVSGBVSRGCGRCVGGVGGSRVSSCSGSS	609						
Qy	141	C 141							
Db	610	610							

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 377.075 Seconds

(without alignments)
11036.790 Million cell updates/sec

Title: US-09-807-802a-1_COPY_1_143

Perfect score: 143

Sequence: 1 ttgccacacccctctcgcg.....aactccacactaggggtaaa 143

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_ov:*

7: gb_ov:*

8: gb_ov:*

9: gb_ov:*

10: gb_ov:*

11: gb_ov:*

12: gb_ov:*

13: gb_ov:*

14: gb_ov:*

15: gb_ov:*

16: gb_ov:*

17: gb_ov:*

18: gb_ov:*

19: gb_ov:*

20: gb_ov:*

21: gb_ov:*

22: gb_ov:*

23: gb_ov:*

24: gb_ov:*

25: gb_ov:*

26: gb_ov:*

27: gb_ov:*

28: gb_ov:*

29: gb_ov:*

30: gb_ov:*

31: gb_ov:*

32: gb_ov:*

33: gb_ov:*

34: gb_ov:*

35: gb_ov:*

36: gb_ov:*

37: gb_ov:*

38: gb_ov:*

39: gb_ov:*

40: gb_ov:*

41: gb_ov:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	4718	14	AF063497 Adeno-ass
2	138.8	97.1	4683	14	AF028704 Adeno-ass
3	119	83.2	4718	14	AF063497 Adeno-ass
4	113.2	79.2	4722	14	AF028705 Adeno-ass
5	113.2	79.2	4726	14	AF028705 Adeno-ass
6	112.2	78.5	4722	14	AF028705 Adeno-ass
7	107.4	75.1	46401	14	AF028705 Adeno-ass
8	107.4	75.1	145	14	AF028705 Adeno-ass
9	107.4	75.1	145	14	AF028705 Adeno-ass
10	107.4	75.1	145	14	AF028705 Adeno-ass
11	107.4	75.1	145	14	AF028705 Adeno-ass
12	107.4	75.1	145	14	AF028705 Adeno-ass
13	107.4	75.1	145	14	AF028705 Adeno-ass
14	107.4	75.1	145	14	AF028705 Adeno-ass
15	107.4	75.1	145	14	AF028705 Adeno-ass
16	107.4	75.1	145	14	AF028705 Adeno-ass
17	107.4	75.1	145	14	AF028705 Adeno-ass
18	107.4	75.1	145	14	AF028705 Adeno-ass
19	107.4	75.1	145	14	AF028705 Adeno-ass
20	107.4	75.1	145	14	AF028705 Adeno-ass
21	107.4	75.1	145	14	AF028705 Adeno-ass
22	107.4	75.1	145	14	AF028705 Adeno-ass
23	107.4	75.1	145	14	AF028705 Adeno-ass
24	107.4	75.1	145	14	AF028705 Adeno-ass
25	105.8	72.9	5585	6	AR054109 Sequence
26	104.2	71.7	5585	6	AR054109 Sequence
27	102.6	69.5	145	14	AF028705 Adeno-ass
28	99.4	69.5	145	14	AF028705 Adeno-ass
29	99.4	69.5	145	14	AF028705 Adeno-ass
30	99.4	69.5	145	14	AF028705 Adeno-ass
31	99.4	69.5	145	14	AF028705 Adeno-ass
32	99.4	69.5	145	14	AF028705 Adeno-ass
33	99.4	69.5	145	14	AF028705 Adeno-ass
34	99.4	69.5	145	14	AF028705 Adeno-ass
35	99.4	69.5	145	14	AF028705 Adeno-ass
36	99.4	69.5	145	14	AF028705 Adeno-ass
37	99.4	69.5	145	14	AF028705 Adeno-ass
38	99.4	69.5	145	14	AF028705 Adeno-ass
39	99.4	69.5	145	14	AF028705 Adeno-ass
40	99.4	69.5	145	14	AF028705 Adeno-ass
41	98.8	69.1	4726	14	AF028705 Adeno-ass

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
AF063497	AF063497	Adeno-associated virus 1, complete genome.	AF063497	AF063497.1	GI:4689096	adeno-associated virus 1.	adeno-associated virus 1.	1 (bases 1 to 4718)	Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and Wilson, J.M.	Gene therapy vectors based on adeno-associated virus type 1	J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE 99214338
PUBMED 10196295
REFERENCE 2 (bases 1 to 4718)
AUTHORS Xiao, M. and Wilson, J.M.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601
Spruce Street, Philadelphia, PA 19104, USA
FEATURES
SOURCE
1. 4718
/organism="adeno-associated virus 1"
/db_xref="taxon:85106"
/note="AAV1"
335..2206
/codon_start=1
/product="nonstructural protein"
/protein_id="AAD27758.1"
/db_xref="GI:4689098"
/translation="MPGFYELIVKPSDLDEHLPGISDSFVSNVAEKEMELPPDSMD
LNIIEQAPLTVAEKLDPLVQMRVSKAPALFPVQFEGESYFHLHILVETGVKS
MVLGRFLSQIRDLVQTIYRGIEPTLPNMFVAHDTLHVSQOEKKNELNPSDPAVRS
TOPELQWAMTMEVYISACINLAERKRLVQHLTHVOTOEONKNEANPSDPAVRS
KTSARVYELVGMVLDRCITSEKQIODEQASYISFVNASRSQIKALNDACKIMAL
TKSAPDYLVPAPADIKTNRIYRIELNGEYPAVSVPLGMAQKRGKRTIWLFG
PATGKTNIAEALAHAVPEGCVMNTNEFPNDQVCKXVIMWEGMGTAKVESAKA
ILGSKRVDOCKCKSAQIDPTPIVISTNMCAVDGNSSTTEHQOPLQDRMEKEL
TRLIEHDFGKVTQKVEKFEPRMAQDHYTEVAHEFYRKSGANKRPAPDDAKSEPKA
CPSEADSTSDAEGAPVDPAFRYONKCSRHAGLQMLPKCTCERNONFNICTHGT
RDSECFPGVSESOPVVRKRYRKLCAIHHLGRABEIASCADLVNVDLDCVSEQ"
2223..4433
/codon_start=1
/product="capsid protein"
/protein_id="AAD27757.1"
/db_xref="GI:4689097"
/translation="MAADGYLPMLDNLSEGIKRWMDLKPGAKPRKANOQKDDRG
LVLEKYLDPFNLGDFPVNADAAALHEDKAYDOOLKAGDPIYLRVHADAEFOE
RLQEDTSFGNLAGAVFOAKRVLLEPLGVEEGAKTAPGKRPEVQSPQSDSSGIG
KTQOPAKRLNFGQTDSESVDPQPLGEPATPAVGPPTMASGGGAPMADNNEGA
DGVNAGSNHNCSTWLGDRVITSTRMALPTYNHLKYQISSASTGASNDNHYFY
STPWGYPDFNRFHCHSPRDMORLINNNGFRKRLNFKLNIQVEVTTNDGVTIA
NMLSTVOVSDSEYOLPYVLGSAHOGCLPPADVPMIPOYGYLTINNGSQAVERSS
PYCLEYFSPQMLRTGNPFSTYEEVFPSSIAHSGSLDLNKLPLIDQLYLYNRQK
NQGSAONKDLFERSGPAGMSVOPKMWLPQPCYRQORVAKTKTDNNNSNFTWLGAK
YINLNGRESIINPTGAMASHKDEKFPMSGVMIKGESAGASNTALDNMTDEBEI
KATNPATERFGTVAVNFOSSSTDPATGAVHMGALPGMWODRDVYLQPIWAKIPH
TDGHFHSPLMGFGFLKHPPOILLINTVPANPAPFASATKPSFITVSTQGVSE
IEMLEOKENSKRNPEVOYTSNKAANDVETVNDNGLYTERPRTIGRYLTRPL"

BASE COUNT 1121 a 1393 c 1273 g 931 t
ORIGIN
Query Match 100.0%; Score 143; DB 14; Length 4718;
Best Local Similarity 100.0%; Pred. No. 5,7e-23;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTCGGTGGGAGCTTGGCGGACCAAGGTCGCG 60
DB 1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTCGGTGGGAGCTTGGCGGACCAAGGTCGCG 60
QY 61 AGAGCGGAGAGCTCTGCTGCGCGCGCGCGCGCGGACGAGCGAGCGGCGGAGAGGAGTG 120
DB 61 AGAGCGGAGAGCTCTGCTGCGCGCGCGCGCGCGGACGAGCGAGCGGAGAGGAGTG 120
QY 121 GGCACTCCATCCTAGGGGTA 143
DB 121 GGCACTCCATCCTAGGGGTA 143

RESULT 2
AF028704/c 4683 bp DNA linear VRL 12-JAN-1998
LOCUS AF028704
DEFINITION Adeno-associated virus 6, complete genome.
ACCESSION AF028704
VERSION AF028704.1 GI:2766605
KEYWORDS

SOURCE adeno-associated virus 6.
ORGANISM adeno-associated virus 6
REFERENCE 1 (bases 1 to 4683)
AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus
JOURNAL Infectious clones and vectors derived from adeno-associated virus
J. Virol. 72 (1), 309-319 (1998)
MEDLINE 98080418
PUBMED 9420229
REFERENCE 2 (bases 1 to 4683)
AUTHORS Rutledge, E.A. and Russell, D.W.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box
357720, Seattle, WA 98195, USA
FEATURES
SOURCE
1. 4683
/organism="adeno-associated virus 6"
/db_xref="taxon:68558"
/note="AAV6"
320..2191
/codon_start=1
/evidence="not_experimental"
/product="nonstructural protein Rep78"
/protein_id="AAB95449.1"
/db_xref="GI:2766606"
/translation="MPGFYELIVKPSDLDEHLPGISDSFVSNVAEKEMELPPDSMD
LNIIEQAPLTVAEKLDPLVQMRVSKAPALFPVQFEGESYFHLHILVETGVKS
MVLGRFLSQIRDLVQTIYRGIEPTLPNMFVAHDTLHVSQOEKKNELNPSDPAVRS
TOPELQWAMTMEVYISACINLAERKRLVQHLTHVOTOEONKNEANPSDPAVRS
KTSARVYELVGMVLDRCITSEKQIODEQASYISFVNASRSQIKALNDACKIMAL
TKSAPDYLVPAPADIKTNRIYRIELNGEYPAVSVPLGMAQKRGKRTIWLFG
PATGKTNIAEALAHAVPEGCVMNTNEFPNDQVCKXVIMWEGMGTAKVESAKA
ILGSKRVDOCKCKSAQIDPTPIVISTNMCAVDGNSSTTEHQOPLQDRMEKEL
TRLIEHDFGKVTQKVEKFEPRMAQDHYTEVAHEFYRKSGANKRPAPDDAKSEPKA
CPSEADSTSDAEGAPVDPAFRYONKCSRHAGLQMLPKCTCERNONFNICTHGT
RDSECFPGVSESOPVVRKRYRKLCAIHHLGRABEIASCADLVNVDLDCVSEQ"
2208..4418
/codon_start=1
/evidence="not_experimental"
/product="capsid protein VP1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translation="MAADGYLPMLDNLSEGIKRWMDLKPGAKPRKANOQKDDRG
LVLEKYLDPFNLGDFPVNADAAALHEDKAYDOOLKAGDPIYLRVHADAEFOE
RLQEDTSFGNLAGAVFOAKRVLLEPLGVEEGAKTAPGKRPEVQSPQSDSSGIG
KTQOPAKRLNFGQTDSESVDPQPLGEPATPAVGPPTMASGGGAPMADNNEGA
DGVNAGSNHNCSTWLGDRVITSTRMALPTYNHLKYQISSASTGASNDNHYFY
STPWGYPDFNRFHCHSPRDMORLINNNGFRKRLNFKLNIQVEVTTNDGVTIA
NMLSTVOVSDSEYOLPYVLGSAHOGCLPPADVPMIPOYGYLTINNGSQAVERSS
PYCLEYFSPQMLRTGNPFSTYEEVFPSSIAHSGSLDLNKLPLIDQLYLYNRQK
NQGSAONKDLFERSGPAGMSVOPKMWLPQPCYRQORVAKTKTDNNNSNFTWLGAK
YINLNGRESIINPTGAMASHKDEKFPMSGVMIKGESAGASNTALDNMTDEBEI
KATNPATERFGTVAVNFOSSSTDPATGAVHMGALPGMWODRDVYLQPIWAKIPH
TDGHFHSPLMGFGFLKHPPOILLINTVPANPAPFASATKPSFITVSTQGVSE
IEMLEOKENSKRNPEVOYTSNKAANDVETVNDNGLYTERPRTIGRYLTRPL"

BASE COUNT 1114 a 1363 c 1277 g 929 t
ORIGIN
Query Match 97.1%; Score 138.8; DB 14; Length 4683;
Best Local Similarity 98.6%; Pred. No. 5.1e-22;
Matches 140; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTCGGTGGGAGCTTGGCGGACCAAGGTCGCG 60
DB 4683 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTCGGTGGGAGCTTGGCGGACCAAGGTCGCG 4624
QY 61 AGAGCGGAGAGCTCTGCTGCGCGCGCGCGCGGACGAGCGAGCGGCGGAGAGGAGTG 120
DB 4623 AGAGCGGAGAGCTCTGCTGCGCGCGCGCGCGGACGAGCGAGCGGCGGAGAGGAGTG 4564
QY 121 GGCACTCCATCCTAGGGGTA 142

BASE COUNT	1204 a	1297 c	1243 g	978 t
Query Match	79.2%	Score 113.2;	DB 14;	Length 4722;
Best Local Similarity	87.3%	Fred. No. 3.2e-16;		
Matches 124;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;
Qy	1	TTGCCCACTCCCTCTCTCGCGGCTCGCTCGCTCGGTGGGGGCTCGCGAGCAAAAGTCCGC	60	
Db	4722	TTGGCCACTCCCTCTCTCGCGGCTCGCTCGCTCGGTGGGGGCTCGCGAGCAAAAGTCCGC	4663	
Qy	61	AGACGGCAGAGCTTGTCTTCCCGGCCCCACCGAGCGAGCGCGGAGAGAGAGTGTG	120	
Db	4662	AGACGGCAGAGCTTGTCTTCCCGGCCCCACCGAGCGAGCGCGGAGAGAGAGTGTG	4603	
Qy	121	GGCAACTCCATCATCAGAGGATG	142	
Db	4602	GGCAACTCCATCATCAGAGGATG	4581	
RESULT 5	AVU48704	4726 bp	DNA	linear
LOCUS	AVU48704			VRL-15-JUL-1996
DEFINITION	Adeno-associated virus 3 nonstructural protein and capsid protein			
ACCESSION	U48704			
VERSION	U48704.1			
KEYWORDS	GI:1408467			
SOURCE	Adeno-associated virus 3 strain=JH.			
ORGANISM	adeno-associated virus 3			
REFERENCE	1 (bases 1 to 4726)			
AUTHORS	Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.			
TITLE	Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3			
JOURNAL	Virology 221 (1), 208-217 (1996)			
MEDLINE	96266430			
PUBMED	8661429			
REFERENCE	2 (bases 1 to 4726)			
AUTHORS	Muramatsu, S. and Brown, K.E.			
TITLE	Direct Submission			
JOURNAL	Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch, NHLBI/NIH, 9000 Rockville, MD 20892, USA			
FEATURES	Location/Qualifiers			
Source	1..4726			
	/organism="adeno-associated virus 3"			
	/strain="JH"			
	/db_xref="taxon:46350"			
	318..2192			
	/codon_start=1			
	/product="nonstructural protein"			
	/protein_id="AAC55048.1"			
	/db_xref="GI:1408468"			
	/translation="MPGEYELVKVPSDLDERLPGISNFPVNWAEKEWDVPSDDMDPNLEQAPLVAEKLQREFLVEMRWVSAPPALFPVQEKETVFLHVLLETIGVKSMMVRYVYSQIEKLVTIRYGVPEQLPMFVATKRNAGKRVDCYIPNVLPRKTQPELOAMVTNMDIYLSACNLIAERKRLVAOHLTHVSTOENKQENPNDAVYSIKTSRAMELVGMIVLDRGITSEKOWTQEOBQAYISGNASNGRSOIKALDASKIMSITKTRADYVGSNPEDITKRNITQYLEANGDPQYAAVFGMAQKQKGTNTLWLGSPATGKTNIAEALIAHVPFGVQWNTNFPNDVDMVIMWBEKKTAKVASEFELILGSKVAVDCKKSSAQIEPTPIVTSNTMCAVDNSTTFEHLQPLQURMEFEETRLLDHGKQVTKQEVKDFPMASSHVDVAHEFVVRGAKKRPASNDADVSPKKBKCTSLAQPPTSDAEAPADYADRYQNKSHVGMNMLFPCKTCERNQOISNCFTHGQRDCEGEFPMSSESQPVSVYKKTQYGLCPHILHGAPEIACSDLANVDDDCVSEG"			
	2209..4419			
	/codon_start=1			
	/product="capsid protein"			
	/protein_id="AAC55049.1"			
	/db_xref="GI:1408469"			
	/translation="MAADGYLPMDEMDLSEGIRHMAALKPGVPOPKAQQOHOHNRGGLVBPRTKYLGRNGIDKCEPYNVEDAALAEHDKAVDQLRKGDNPYLYKNALAEPOELVBOETSTSGNIGRAVFOAKRILEPLVEBAATLGAAGKGAVDQSEDFSSGAGV"			

BASE COUNT	1200 a	1295 c	1251 g	980 t
Query Match	79.2%;	Score 113.2;	DB 14;	Length 4726;
Best Local Similarity	87.3%;	Pred. No. 3.2e-16;		
Matches 124;	Conservative	0;	Mismatches 18;	Indels 0;
Gaps	0;			
Db	1	TTGCCCACTCCCTCTGCGGCGCTCGTGGTGGTGGGCGCTTGCAGCAACCAAGTCCGC	60	
Qy	1	TTGCCCACTCCCTCTGCGGCGCTCGTGGTGGTGGGCGCTTGCAGCAACCAAGTCCGC	60	
Db	61	AGACGACGAGAGCTTGTCTGTGCGCGCCGCCACCGACGACGACGCGCGCAGAGAGAGTGTG	120	
Qy	61	AGACGACGAGAGCTTGTCTGTGCGCGCCGCCACCGACGACGACGCGCGCAGAGAGAGTGTG	120	
Db	121	GCGAATCTTCATCTACGAGGTTA	142	
Qy	121	GCGAATCTTCATCTACGAGGTTA	142	
Db	121	GCGAATCTTCATCTACGAGGTTA	142	
RESULT 6				
AF028705	AF028705	4722 bp	DNA	linear
LOCUS	Adeno-associated virus 3B, complete genome.			VRL 12-JAN-1998
DEFINITION	Adeno-associated virus 3B, complete genome.			
ACCESSION	AF028705			
VERSION	AF028705.1	GI:2766608		
KEYWORDS				
SOURCE	adeno-associated virus 3B.			
ORGANISM	adeno-associated virus 3B.			
ORGANISM	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
REFERENCE	1 (bases 1 to 4722)			
AUTHORS	Rutledge, E.A., Halbert, C.L. and Russell, D.W.			
TITLE	Infectious clones and vectors derived from adeno-associated virus			
JOURNAL	(AAV) serotypes other than AAV type 2			
PUBMED	J. Virol. 72 (1), 309-319 (1998)			
REFERENCE	2 (bases 1 to 4722)			
AUTHORS	Rutledge, E.A. and Russell, D.W.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720,			
PUBMED	Seattle, WA 98195, USA			
FEATURES	Location/Qualifiers			
Source	1..4722			
	/organism="adeno-associated virus 3B"			
	/db_xref="taxon:68742"			
	/note="AAV3B"			
	317..2191			
	/product="nonstructural protein Rep78"			
	/codon_start=1			
	/protein_id="AAB5451.1"			
	/db_xref="GI:2766609"			
	/translation="MPGFYEIVLVKPSDLDEHLPGISNPFVNVAAKEWELPPDSMDP			
	PNLEQALPIVAEKLQREPLVEMRWVSKAEALFVQPEKGETYFHLVLIETIGVSKS			
	MMVGRVYSQIEKLVRIYRGVEPOLNMFVAVTKTRGAGAGNVDDCYIPNLLPKR			
	TOPLQIAWMTNMOYSACINLAERKLVNQHILTVASOTGONKNNKPNNSDAVIRIS			
	KTSAPRIELVGMVLYDRGITSSEKQWIDQASVYISFMAASRSRQIKALDNASKIMSL			
	TKTAPDLVGSNPEDDITKRIYQIYDLNGDPYASVFGVHQAQKFRKRNITWLG			
	PATGSKNIAEALAAHAVPYGCVMWNTNEFPNDQVKNVIWEEGKMTAAVVSARAI			
	ILGSKRVVDQKCSQAIEPTPIYVSNNTMCVIDGNSITTEPHOPLDQRMKFEI			
	TRRLDPAVEQYKQEVQEPFRMSADHTDVAAHEFYVKGAKKAPASNDADVSPKQ			
	CTSLAQTTSDAEAPADYARVYQNGKGRHGMNMLPPCKTCERMMOISNCFTHGQR			
	DGCECFPMESQSPVSVYKKTKYKLCRPHIHLIGRAELNAGCADLANVDLDDCVSQC			

CDS

```

2208 . . .4418
/codon_start=1
/product="capsid protein VP1"
/protein_id="AA095452.1"
/db_xref="gi:2766610"
/translation="MAADGVLPWLEEDNTSEGIREWMAALKEPVDPKANOQHONRRG
LVLPVETGKGLNGNDLGEVSEENADAALEHDKAVDQQLKADNPLYLKYNAADPEDE
LRLQEDTSGKGLGAVYPAQAKRLILEPGLIVEPAKAPKAPVDSQEPDSSGVGVA
KSGQAPAKRLNFOETGDSVSDVPTPLGAPKAPVSLGSLNMGAGGAPMADNNEGVA
KGVSSGNHMDLQEDTGDSDVDEKPTITSRMALPYNNHLKQISQSGAGAPMADNNEGVA
TPKVPYEPNRPKHSFPRVDMORLLNNMGKPKKLSFKLFINQVEYVQNGGTTNLA
NLSTVQVPTFSDTEQLPVEYVGSAGHCGCLPPFPADVPMVQYQYVLYLNGSQAVRSS
YCLEFSPQMLRTNNQFSTVYGSADGDFHSHASQSDLRNLNPLIDQTLVLRNQLD
THLNGROSLRFLQAPQAGQSMISQARNLWLPQPCYRQRLSTKLAANDNNNPNWPAAS
HLNGROSLVNPGRGAPMAHQDEKPEPMGNLIIPKEKETTSLANNLNNLWNPITDEE
RTTPVATVEQGYANMLQGSNNLSTANLPTFVYLDQGLAPGVWQDDRVYVLYQGIYAKI
TDGHHFPLDLGGGGLGAPPEQPLNIKTPVPAVPPTTFSPAKFASFIYQSGVQSVSE
IWELEOKNSKRNMPETQYTSNNKNSVNDVFTVDNNGVSEBRPIGRTYLRRLRNL"
a      1297 c      1243 g      978 t

```

BASE COUNT
ORIGIN

Query Match	78.5%;	Score 112.2;	DB 14;	Length 47722;
Best Local Similarity	87.2%;	Pred. No. 5.4e-16;		
Matches 123;	Conservative	0;	Mismatches 18;	Indels 0;
				Gaps 0;

Qy 2 TGCCACTCCCTCTGCGCGCTCGCTCGGTTGGGGCTGCGGACCAAGGTGCGCA 61

Db 1 TGCCACTCCCTCTATGCGCACTCGCTCGGTGGGGCTTGCGGACCAAGGTGCGCA 60

Dy 62 GACGGCAGACTCTGCTCGCCGCCACCAGCGAGACGCAGAGGGAGTGG 121
Db 61 GACGGACGTCTTTGACGTCGCCGCCACCGAGCGAGCGAATGCCATTAGAGGGAGTGG 120

Oy		122	GCAACTCCATCACAAGGGTA	142
Dd		121	CCACTTCCATCATTAGAAGTA	141

RESULT 7			
A46401			
LOCUS	A46401	145 bp	DNA
DEFINITION	Sequence 4 from Patent WO9523867.		PAT 07-MAR-1997

VERSION	A46401.1	GI:2300602
KEYWORDS		
SOURCE	unidentified.	
ORGANISM	unidentified	
	unclassified.	

REFERENCE	141 (bases 1 to 145)
AUTHORS	Denefle, P., Latta, M., Perricaudet, M. and Vigme, E.
TITLE	INTERACTIVE RECOMBINANT ADENOVIROUSES, PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF
JOURNAL	Patent: WO 9523867-A 4 08-SEP-1995; RHONE POULENC RORER SA (FR)
COMMENT	Other publication AU 1852595 950918 Other publication FR 2716893 950908.
COMMENT	

FEATURES	Location/Qualifiers
source	1. .145

BASE COUNT	21	a	52	c	49	g	23	t
ORIGIN	/organism="unidentified" /db_xref="taxon:32644"							

Query Match	75.1%;	Score 107.4;	DB 6;	Length 145;
Best Local Similarity	85.1%;	Pred. No. 1.3e-14;		
Matches 120;	Conservative	0;	Mismatches 21;	Indels 0;
			Gaps	0

Qy 1 TTGCCCATCCTCTCTGTGGCGCCTGCTCGCTCGGTGGGGCTTGGGGAACAAAGTTCGCG 60
Db 1 TTGGCCACTCCTCTCTGTGGCGCCTGCTCGCTCTGACTGAGCGCGGGGAGCAAAAGTTCGCG 60
Qy 61 AGACGGCAGAGCTCTGCTCTGCGCGGCCCCAACCGAGCGAGCGCGCAGAGAGGAGTGC 120

Dib

61 CGACGCCCGGCTTTTGTCCCGGGCGGCCTCAGTAGCGAAGCACGCGCAGAAGAGGAGTG 120
121 GGCACCTCCATCATCTNAGGGT 141
122 | | | | |
121 GCCACTCCATCATCTNAGGGT 141

RESULT 8
AA2LTR1

LOCUS	A212T1	145 bp	DNA	linear	VRL 27-APR-1993
DEFINITION	Adeno-associated virus 2	left terminal	sequence.		
ACCESSION	K01624				
VERSION	K01624.1	GI:209623			
KEYWORDS	replication; terminal repeat.				
SEGMENT	1 of 2				
SOURCE	Adeno-associated virus 2H DNA, (clone pSM620 [2]), from KB or HeLa				

ORGANISM	REFERENCE
adeno-associated virus 2H	1 (bases 1 to 145)
Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.	

TITLE Nucleotide sequence of the inverted terminal repetition in
adeno-associated virus DNA
J. Virol. 34 (2), 402-409 (1980)
JOURNAL
0022-5381

PUBMED	6246271
REFERENCE	2 (bases 1 to 145)
AUTHORS	Lefebvre, R.B., Riva, S. and Berns

Virus DNA replication
Mol. Cell. Biol. 4 (7), 1416-1419 (1984)
JOURNAL
MEDLINE
85061247
5704040
0706040

COMMENT
Both [1] and [2] present the opposite strand from the one presented here. The focus of both papers is the method of replication of the virus. [1] notes that the initial ϵ is present only 30% of the

heterogeneity which can be explained by assuming that the terminal 125 bases, which form an imperfect palindrome, are replaced by their inverted complement during replication. [2] found that deletion of the 9 terminal bases on the right and the 113 terminal

Further deletion of an 11-base symmetrical sequence (bases 89 to 99) in the right terminal repetition inhibits DNA replication. Substitution of either an 8-base (cagatctg) or 12-base (cgcgagtcgcg) symmetrical sequence unrelated to the original 11-base sequence restores DNA replication. All of this can be

```

FEATURES
  source
    1. .145
      /organism="adeno-associated virus 2H"
      /db_xref="taxon:10805"
  21 a
    52 c
    49 g
    23 t

```

ORIGIN 2 bases upstream of HaeIII site

Query Match	75.1%;	Score 107.4;	DB 14;	Length 145;
Best Local Similarity	85.1%;	Pred. No. 1.3e-14;		
Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

Db 1 TTGGCACTCCCTCTGCGCGCTGCTGCTCACTGAGGCGGGGACCAAGGTGCC 60

D7 AGATGGCAGAACCTTCCTCTGCTCCCGGCACCAAGCCAGGCCAGTGTGAGAGAGAAAG 120
| | | | | | | | | | | | | | | | | | | | | |
D8 61 CGAGGCCCGGCTTTGGCCCGGGCGGCTCACTGAGCGAGCGAGCGGCAGAGAGGAAGTG 120
| | | | | | | | | | | | | | | | | | | | | |
QY 121 GGCAACTCATCACTAAGGGT 141
| | | | | | | | | | | | | | | |

Db 121 GCCAATCTCATCTAGGGGT 141

RESULT 9
AA2REPORT 145 bp ss-DNA linear VRL 27-APR-1993
LOCUS Adeno-associated virus origin of replication (genome 3' terminus).
DEFINITION M10681
ACCESSION M10681
VERSION M10681.1 GI:209626
KEYWORDS
SOURCE Adeno associated virus 2H (AAV2 H) DNA.
ORGANISM adeno-associated virus 2H
REFERENCE 1 (bases 1 to 145)
AUTHORS Berns, K.I., Hauswirth, W.W., Pffe, K.H. and Luby, E.
TITLE Adeno-associated virus DNA replication
JOURNAL Cold Spring Harb. Symp. Quant. Biol. 43 Pt 2, 781-787 (1979)
MEDLINE 80023388
PUBMED 226321
FEATURES
source Location/Qualifiers
1..145
/organism="adeno-associated virus 2H"
/db_xref="taxon:10805"
BASE COUNT 23 a 49 c 52 g 21 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 145;
Best Local Similarity 85.1%; Pred. No. 1.3e-14;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGC 60
Db 145 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCGCTCACTAGGCGCGGCAAGGTCGCC 86

Qy 61 AGAGCGAGAGCTCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 85 CGAGCGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 26

Qy 121 GGCACTCATCTAGGGGT 141
Db 25 GCCAATCTCATCTAGGGGT 5

RESULT 10
AR034135 165 bp DNA linear PAT 29-SEP-1999
LOCUS AR034135
DEFINITION Sequence 1 from patent US 5869305.
ACCESSION AR034135
VERSION AR034135.1 GI:5949740
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 165)
AUTHORS Samulski, R.Ude. and Xiao, X.
TITLE Recombinant viral vector system
JOURNAL Patent: US 5869305-A 1 09-FEB-1999;
FEATURES
source Location/Qualifiers
1..165
/organism="unknown"
BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.3e-14;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGC 60
Db 21 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTAGGCGCGGCAAGGTCGCC 80

Qy 61 AGAGCGAGAGCTCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 11 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGCC 120

Db 81 CGAGCGCGGCGCTTGGCGGCGGCGGCTCAGTAGAGGAGCGGCGGCGGAGGAGTG 140

Qy 121 GGCACTCATCTAGGGGT 141
Db 141 GCCAATCTCATCTAGGGGT 161

RESULT 11
AX106702 165 bp DNA linear PAT 30-APR-2001
LOCUS AX106702
DEFINITION Sequence 7 from Patent WO0125465.
ACCESSION AX106702
VERSION AX106702.1 GI:13922363
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 165)
AUTHORS Engelhardt, J.F., Dongsheng, D. and Ziyang, Y.
TITLE Adeno-associated viruses and uses thereof
JOURNAL Patent: WO 0125465-A 7 12-APR-2001;
University of Iowa Research Foundation (US);
(US); Dongsheng, Duan (US); Ziyang, Yan (US)
FEATURES
source Location/Qualifiers
1..165
/organism="unidentified"
/db_xref="taxon:32644"
/note="SEQ ID NO:1 of U.S. Patent No. 5,478,745"
BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.3e-14;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGC 60
Db 21 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTAGGCGCGGCAAGGTCGCC 80

Qy 61 AGAGCGAGAGCTCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 81 CGAGCGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 140

Qy 121 GGCACTCATCTAGGGGT 141
Db 141 GCCAATCTCATCTAGGGGT 161

RESULT 12
I16806 165 bp DNA linear PAT 03-APR-1996
LOCUS I16806
DEFINITION Sequence 1 from patent US 5478745.
ACCESSION I16806
VERSION I16806.1 GI:1251714
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 165)
AUTHORS Samulski, R.U. and Xiao, X.
TITLE Recombinant viral vector system
JOURNAL Patent: US 5478745-A 1 26-DEC-1995;
FEATURES
source Location/Qualifiers
1..165
/organism="unknown"
BASE COUNT 27 a 56 c 56 g 26 t
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 165;
Best Local Similarity 85.1%; Pred. No. 1.3e-14;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGC 60
Db 1 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGGCTTGCAGACCAAGGTCGCC 60

Db	21	TTGGCCACTCCTCTCTCTGCGGCTGGCTGGCTCACTAGAGCCGGGCGACCAAAGTCCGC	80
Qy	61	AGACGGCAGAGCTCTGCTCTGCGGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTG	120
Db	81	CGACGCCCGGGGCTTTGCCCGGGCGGCTCACTGAGCGAGCGCGCAGAGAGGAGTG	140
Qy	121	GGCAACTCCATCATCTAGGGGT	141
Db	141	GCCAACTCCATCATCTAGGGGT	161
RESULT 13			
LOCUS	AX106701	272 bp	DNA
DEFINITION	Sequence 6 from Patent WO0125465.		linear
ACCESSION	AX106701		
VERSION	AX106701.1	GI:13922362	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 272)		
AUTHORS	Engelhardt, J. F., Dongsheng, D. and Ziyang, Y.		
TITLE	Adeno-associated viruses and uses thereof		
JOURNAL	Patent: WO 0125465-A 6 12-APR-2001;		
	University of Iowa Research Foundation (US) ; Engelhardt, John F.		
	(US) ; Dongsheng, Duan (US) ; Ziyang, Yan (US)		
FEATURES	Location/Qualifiers		
source	1..272		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
	/note="NAV circular intermediate, clone p1202"		
BASE COUNT	58 a	78 c	78 g
ORIGIN			58 t
Query Match	75.1%;	Score 107.4;	DB 6;
Best Local Similarity	85.1%;	Pred. No. 1.2e-14;	
Matches 120;	Conservative 0;	Mismatches 21;	Indels 0;
Gaps 0;			
Qy	1	TTGCCCACTCCCTCTCTGCGGCTGGCTGGCTGGGCTGGGCGACCAAAGTCCGC	60
Db	69	TTGGCCACTCCTCTCTCTGCGGCTGGCTGGCTCACTGAGGCGCGACCAAAGTCCGC	128
Qy	61	AGACGGCAGAGCTCTGCTCTGCGGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTG	120
Db	129	CGACGCCCGGGGCTTTGCTGCGCGGCTCACTGAGCGAGCGCGCAGAGAGGAGTG	188
Qy	121	GGCAACTCCATCATCTAGGGGT	141
Db	189	GCCAACTCCATCATCTAGGGGT	209
RESULT 14			
LOCUS	AA2LEFT	2116 bp	DNA
DEFINITION	adeno-associated virus 2 left half 45% of genome.		linear
ACCESSION	J01902		
VERSION	J01902.1	GI:209622	
KEYWORDS			
SOURCE	adeno-associated virus 2 from human hela cells.		
ORGANISM	adeno-associated virus 2		
REFERENCE	1 (bases 1 to 2116)		
AUTHORS	Lusby, E. W. and Berns, K. I.		
TITLE	Mapping of the 5' termini of two adeno-associated virus 2 RNAs in		
	the left half of the genome		
JOURNAL	J. Virol. 41 (2), 518-526 (1982)		
MEDLINE	82192580		
PUBMED	6281463		
FEATURES	Location/Qualifiers		
source	1..2116		
	/organism="adeno-associated virus 2"		

BASE COUNT	526 a	-539 c	615 g	436 t
ORIGIN	/db_xref="taxon:10804"			
Query Match	75.1%; Score 107.4; DB 14; Length 2116;			
Best Local Similarity	85.1%; Pred. No. 7.7e-15;			
Matches 120; Conservative	0; Mismatches 21; Indels 0; Gaps 0;			
Oy	1	TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCGGTGGGGCTTGCGAGACCAAGTCCGC	60	
Db	1	TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCACTAGAGGCGGGGACCAAGTCCGC	60	
Oy	61	AGAGCGGAGAGCTGTGCTTTCGCGCGGCCCCACCGAGCGAGCGAGCGCCAGAGAGGAGTG	120	
Db	61	CGAGCGCCCGGCGCTTTGGCCCGGCGCGCTCACTAGAGCGAGCGAGCGCCAGAGAGGAGTG	120	
Oy	121	GGCAACTCCATCACTAGAGGGT 141		
Db	121	GGCAACTCCATCACTAGAGGGT 141		
RESULT 15				
AX135805	4675 bp	DNA	Linear	PAT 29-MAY-2001
LOCUS				
DEFINITION	Sequence 5 from Patent WO0132711.			
ACCESSION	AX135805			
VERSION	AX135805.1 GI:14272040			
KEYWORDS	adeno-associated virus 2.			
SOURCE	adeno-associated virus 2.			
ORGANISM	Virus; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
REFERENCE	1 (bases 1 to 4675)			
AUTHORS	Hermomat, P.L.			
TITLE	Repo-associated virus aav rep78 major regulatory protein, mutants thereof and uses thereof			
JOURNAL	Patent: WO 0132711-A 5 10-MAY-2001.			
FEATURES	THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS (US)			
source	Location/Qualifiers			
	1..4675			
	/organism="adeno-associated virus 2"			
	/db_xref="taxon:10804"			
CDS	321..2186			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAC39573.1"			
	/db_xref="GI:14272041"			
	translation="MGGFPEIVIVKPSDLDGHLPGTSDSFVNVAEKEMLPDPSMDMLNLEQAPLVAEKLDQDFLEBRVRVSKAPBALFFVDFPEKGESTFPHNVLVETTVKASMVLGRFLSQIREKLIQRIYRGIEPTLPENNAVATKTRNGAGGKVVDECIENYLLPKTQBELQWAMTNMEIYGLASCLNTERKLVQNHQITHVSQTOENKONPNSDAVISIKETARVWELVGMVLDKGIITSEKMOIQDDASYISFNPAASRSQIKALNAGIMELTKPAPYLVGCGOPVEDISNRIYKILLEINVDPOVASVFLGMAATKFKGRNTIWLFGPTATGKTNIMAEIAHTVPPFGCVNMTNENPNPDCVDKMTKIMNEGMTAKVVSAAKAILGGSKRVDQCKSSAQIDPPTVITVSTNMCAVIDGNSITFEHQPLDQRMFKFELTRRLDHPSTGTOEVDVDFPRMAKDHYEHEFHYVKGAGAKRPAISDADISPPKVRRESYADHPTSDAEASINVDARYONKCSRHGMMIMLFPCHOCERMNQNSNICPTHGQKRDCELPFVSESOPSVYVKAYOYLCTVYHNIMKVPDACTACDVLVNDLDICIFEQ"			
BASE COUNT	1198 a	1262 c	1251 g	964 t
ORIGIN				
Query Match	75.1%; Score 107.4; DB 6; Length 4675;			
Best Local Similarity	85.1%; Pred. No. 6.6e-15;			
Matches 120; Conservative	0; Mismatches 21; Indels 0; Gaps 0;			
Oy	1	TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCGGTGGGGCTTGCGAGACCAAGTCCGC	60	
Db	1	TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCACTAGAGCGGGGCAACAAAGTCCGC	60	
Oy	61	AGAGCGGAGAGCTGTGCTTTCGCGGCCCCACCGAGCGAGCGAGCGCCAGAGAGGAGTG	120	
Db	61	CGAGCGCCCGGCGCTTTGGCCCGGCGCGCTCACTAGAGCGAGCGAGCGCCAGAGAGGAGTG	120	
Oy	121	GGCAACTCCATCACTAGAGGGT 141		
Db	121	GGCAACTCCATCACTAGAGGGT 141		

Mon Jul 21 10:00:52 2003

us-09-807-802a-1_copy_1_143.rge

Page 8

Db 121 GCCAATCCATCCTAGGGT 141

Search completed: July 20, 2003, 13:19:53
Job time : 377.075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:53:46 ; Search time 32.1207 Seconds
(without alignments)
10025.800 Million cell updates/sec

Title: US-09-807-802a-1_COPY_1_143

Perfect score: 143
Sequence: 1 tttccaccctctctgcg.....aactccatcactagggttaa 143

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	143	21	AAD00781 Adeno-associated v
2	143	100.0	4718	21	AAD00772 Adeno-associated v
3	138.8	97.1	4683	22	AAFP3749 AAV6 DNA sequence.
4	119	83.2	4718	22	AAFP3748 Adeno-associated v
5	113.2	79.2	4722	22	AAFP3748 AAV3B DNA sequence
6	112.2	78.5	4722	22	AAFP3748 AAV3B DNA sequence
7	107.4	75.1	145	14	AAQ1448 AAV2 inverted term
8	107.4	75.1	145	16	AAFP3385 Strict inverted te
9	107.4	75.1	145	20	AAK34295 Adeno-associated v

10	107.4	75.1	165	15	AAQ6769 Double-D ITR termi
11	107.4	75.1	165	18	AAFP9462 Adeno associated v
12	107.4	75.1	165	24	AAFP3535 Inverted terminal
13	107.4	75.1	272	21	AAZ47166 Head-to-tail ITR s
14	107.4	75.1	272	22	AAFP3534 AAV circular inter
15	107.4	75.1	955	24	AAFP37254 Adeno-associated v
16	107.4	75.1	955	24	AAFP37254 Adeno-associated v
17	107.4	75.1	987	24	AAFP37261 Adeno-associated v
18	107.4	75.1	987	24	AAFP37261 Adeno-associated v
19	107.4	75.1	4414	24	AAFP37260 Adeno-associated v
20	107.4	75.1	4414	24	AAFP37260 Adeno-associated v
21	107.4	75.1	4476	24	AAFP37259 Adeno-associated v
22	107.4	75.1	4476	24	AAFP37259 Adeno-associated v
23	107.4	75.1	4498	24	AAFP37258 Adeno-associated v
24	107.4	75.1	4498	24	AAFP37258 Adeno-associated v
25	107.4	75.1	4675	22	AAH41481 Adeno-associated v
26	107.4	75.1	4675	22	AAFP89311 Nucleotide sequenc
27	107.4	75.1	4675	24	ABA02989 Adeno-associated v
28	107.4	75.1	4679	22	AAI66974 Adeno-associated v
29	107.4	75.1	4679	22	AAFP3750 AAV2 DNA sequence.
30	107.4	75.1	4680	17	AAFP09008 Wild-type adeno-as
31	107.4	75.1	4683	22	AAFP3749 AAV6 DNA sequence.
32	107.4	75.1	4825	24	AAFP37257 Adeno-associated v
33	107.4	75.1	4825	24	AAFP37257 Adeno-associated v
34	107.4	75.1	4848	24	AAFP37263 Adeno-associated v
35	107.4	75.1	4966	24	AAFP37256 Adeno-associated v
36	107.4	75.1	4966	24	AAFP37256 Adeno-associated v
37	107.4	75.1	4980	24	AAFP37262 Adeno-associated v
38	107.4	75.1	4990	24	AAFP37262 Adeno-associated v
39	107.4	75.1	5060	24	AAFP37264 Adeno-associated v
40	107.4	75.1	5060	24	AAFP37264 Adeno-associated v
41	107.4	75.1	5149	24	AAFP37255 Adeno-associated v
42	107.4	75.1	5149	24	AAFP37255 Adeno-associated v
43	107.4	75.1	5322	21	AAZ45928 Nucleotide sequenc
44	107.4	75.1	5322	21	AAZ45928 Nucleotide sequenc
45	107.4	75.1	6142	21	AAZ45932 Nucleotide sequenc

ALIGNMENTS

RESULT 1	
AAD00781	standard; DNA; 143 BP.
ID	AAD00781
XX	08-SEP-2000 (first entry)
XX	Adeno-associated virus serotype 1 5' inverted terminal repeat.
XX	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KM	cap protein; recombinant viral vector; gene delivery; gene therapy;
KW	vaccine; transgene; inverted terminal repeat; ITR; ss.
XX	
OS	Adeno associated virus serotype 1.
XX	
FH	Key
FT	misc_binding
FT	Location/Qualifiers
FT	1..41
FT	/tag= a
FT	/bound_molecy= "binds to nucleotides 85..125"
FT	stem_loop
FT	42..62
FT	/tag= b
FT	stem_loop
FT	64..84
FT	/tag= c
FT	misc_binding
FT	85..125
FT	/tag= d
XX	/bound_molecy= "binds to nucleotides 41..1"
XX	
XX	MO200028061-A2.
XX	
XX	18-MAY-2000.
XX	

PF 02-NOV-1999; 99WO-US25694.
XX 05-NOV-1998; 98US-0107114.
XX (VYPE-) UNIV PENNSYLVANIA.
XX Wilson JM, Xiao W;
XX WPI; 2000-376571/32.
DR
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
XX
PS Example 2; Fig 2; 108pp; English.
XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterized by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is a 5' ITR of AAV-1 DNA which is useful in the
XX production of recombinant viral vector. The ITR forms a T-shaped hairpin
XX structure.
SQ Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;
Query Match 100.0%; Score 143; DB 21; Length 143;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTGCGGAGGAGTCCGC 60
DB 1 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTGCGGAGGAGTCCGC 60
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCTCGCTCGCTGCGGAGGAGTCCGC 120
DB 61 AGACGGCAGAGCTCTGCTCTGCGGCTCGCTCGCTGCGGAGGAGTCCGC 120
QY 121 GGCACCTCCTACTAGGGGTAA 143
DB 121 GGCACCTCCTACTAGGGGTAA 143
RESULT 2
AADD00772
ID AADD00772 standard; DNA; 4718 BP.
XX
AC AADD00772;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 DNA.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; ss.
XX
XX Adeno associated virus serotype 1.
OS
XX
FH Key Location/Qualifiers
FT repeat_unit 1..143
FT /tag= a
FT /label= 5' ITR
FT /note= "inverted terminal repeat which is capable
FT of forming T-shaped hairpin structure"
FT protein_bind 89..110

FT /tag= b
FT /bound_motety= "Rep protein"
FT 124..125
FT /tag= c
FT /note= "Terminal resolute site (TRS)"
FT 219..226
FT /tag= d
FT /bound_motety= "USF"
FT /note= "E box"
FT 236..299
FT /tag= e
FT /label= p5_promoter
FT 237..245
FT /tag= f
FT /bound_motety= "YY1 factor"
FT 270..275
FT /tag= g
FT /label= p5_TATA-Box
FT 299..306
FT /tag= h
FT /note= "YY1/p5 RNA"
FT 335..2206
FT /tag= i
FT /product= "Rep 78"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 335..2272
FT /tag= j
FT /product= "Rep 68"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 1924..2220
FT /tag= k
FT /note= "This region interrupts the coding sequence
FT of Rep 68 and Rep 40"
FT 857..862
FT /tag= l
FT /label= p19_TATA_Box
FT 882..883
FT /tag= m
FT /note= "P19 RNA"
FT 1007..2206
FT /tag= n
FT /product= "Rep 52"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 1007..2272
FT /tag= o
FT /product= "Rep 40"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 1836..1841
FT /tag= p
FT /label= p40_TATA-BOX
FT 1875..1876
FT /tag= q
FT /note= "p40 RNA"
FT 2223..4433
FT /tag= r
FT /product= "VP1 protein"
FT /note= "Capsid protein"
FT 2634..4433
FT /tag= s
FT /product= "VP2 protein"
FT /note= "Capsid protein"
FT /partial
FT 2829..4433
FT /tag= t
FT /product= "VP3 protein"
FT /note= "Capsid protein"
FT 4447..4452
FT polyA_signal

```
FT      repeat_unit      /*tag= u
FT      4576..4718
FT      /*tag= v
FT      /label=3' ITR
FT      /note="Inverted terminal repeat which is capable
FT      of forming T-shaped hairpin structure"
XX      MO200028061-A2.
XX      18-MAY-2000.
XX      02-NOV-1999; 99WO-US25694.
XX      05-NOV-1998; 98US-0107114.
XX      (UNIV PENNSYLVANIA.
XX      Wilson JM, Xiao W;
XX      WPI; 2000-376571/32.
XX      P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
XX      AAY71169.
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host -
XX      Claim 1; Fig 1; 108bp; English.
XX      The present sequence is an adeno-associated virus serotype 1 (AAV-1)
XX      DNA characterised by two inverted terminal repeats (ITR) and open
XX      reading frames for rep and capsid (cap) proteins. The rep reading frame
XX      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX      reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX      coding regions, are useful in production of recombinant viral vectors
XX      for gene delivery. These vectors can be used as gene therapy.
XX      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX      not induce the formation of neutralising antibodies specific to any
XX      serotype of AAV hence is useful for transforming host cells, and in the
XX      preparation of a medicament for the delivery of transgene to a host.
XX      Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
XX      Query Match 100.0%; Score 143; DB 21; Length 4718;
XX      Best Local Similarity 100.0%; Pred. No. 1.2e-28;
XX      Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TTGGCCACTCCTCTCTGCGGCTCGCTCGCTCGGCGGCTGCGGCAAAAGTCCGC 60
DB      1 TTGGCCACTCCTCTCTGCGGCTCGCTCGCTCGGCGGCTGCGGCAAAAGTCCGC 60
QY      61 AGACGGGAGAGCTCTGCTCTGCGGCCCGCCACCGAGCGAGCGCGCCAGAGGAGTG 120
DB      61 AGACGGGAGAGCTCTGCTCTGCGGCCCGCCACCGAGCGAGCGCGCCAGAGGAGTG 120
QY      121 GGCACTCCATCACTAGGGGTAA 143
DB      121 GGCACTCCATCACTAGGGGTAA 143
QY      121 GGCACTCCATCACTAGGGGTAA 143
DB      121 GGCACTCCATCACTAGGGGTAA 143
XX      RESULT 3
XX      AAF23749/c
XX      AAF23749 standard; DNA; 4683 BP.
XX      AAF23749;
XX      28-MAR-2001 (first entry)
XX      AAV6 DNA sequence.
XX      AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX      atherosclerosis; sickle cell anaemia; thalassemia;
XX      blood clotting disorder; diabetes; ss.
```

```
XX      Adeno associated virus.
XX      OS      US6156303-A.
XX      PN      05-DEC-2000.
XX      PD      11-JUN-1997; 97US-0873168.
XX      PF      11-JUN-1997; 97US-0873168.
XX      PR      11-JUN-1997; 97US-0873168.
XX      PA      (UNIW ) UNIV WASHINGTON.
XX      PI      Russell DW, Rutledge EA;
XX      WPI; 2001-060164/07.
XX      Adeno-associated virus serotype 6 and viral vector derived from it for
XX      gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX      syndrome, sickle cell anemia, thalassemia and diabetes -
XX      Claim 1; Fig 1; 50pp; English.
XX      The present invention relates to adeno-associated virus serotypes. The
XX      CC      present sequence is the DNA sequence of one such serotype (AAV6). AAV6
XX      CC      can be used to construct AAV viral vectors for use in gene therapy for a
XX      CC      range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX      CC      sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
XX      CC      The AAV viral vectors have increased transduction efficiency of a
XX      CC      particular host cell as the AAV virion containing the AAV vector genome
XX      CC      can be modified to express a capsid protein of an AAV serotype that
XX      CC      transduces the selected host cell.
XX      Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
XX      Query Match 97.1%; Score 138.8; DB 22; Length 4683;
XX      Best Local Similarity 98.6%; Pred. No. 1.5e-27;
XX      Matches 140; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 TTGGCCACTCCTCTCTGCGGCTCGCTCGCTCGGCGGCTGCGGCAAAAGTCCGC 60
DB      4683 TTGGCCACTCCTCTCTGCGGCTCGCTCGCTCGGCGGCTGCGGCAAAAGTCCGC 4624
QY      61 AGACGGGAGAGCTCTGCTCTGCGGCCCGCCACCGAGCGAGCGCGCCAGAGGAGTG 120
DB      4623 AGACGGGAGAGCTCTGCTCTGCGGCCCGCCACCGAGCGAGCGCGCCATGAGGAGTG 4564
QY      121 GGCACTCCATCACTAGGGGTAA 142
DB      4563 GGCACTCCATCACTAGGGGTAA 4542
XX      RESULT 4
XX      AAD00772/c
XX      AAD00772 standard; DNA; 4718 BP.
XX      AAD00772;
XX      08-SEP-2000 (first entry)
XX      Adeno-associated virus serotype 1 DNA.
XX      Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX      KM      cap protein; recombinant viral vector; gene delivery; gene therapy;
XX      KM      vaccine; transgene; ss.
XX      Adeno associated virus serotype 1.
XX      OS      Key
XX      FH      repeat_unit      Location/Qualifiers
XX      FT      1..143
XX      FT      /*tag= a
XX      FT      /label=5' ITR
XX      FT      /note="Inverted terminal repeat which is capable
```

```
FT      of forming T-shaped hairpin structure"
FT      89..110
FT      /tag= b
FT      /bound_moiety= "Rep protein"
FT      124..125
FT      /tag= c
FT      /note= "Terminal resolve site (TRS)"
FT      219..226
FT      /tag= d
FT      /bound_moiety= "USF"
FT      /note= "E box"
FT      236..299
FT      /tag= e
FT      /label= p5_promoter
FT      237..245
FT      /tag= f
FT      /bound_moiety= "YY1 factor"
FT      270..275
FT      /tag= g
FT      /label= p5_TATA-Box
FT      299..306
FT      /tag= h
FT      /note= "YY1/p5 RNA"
FT      335..2206
FT      /tag= i
FT      /product= "Rep 78"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      335..2272
FT      /tag= j
FT      /product= "Rep 68"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1924..2220
FT      /tag= k
FT      /note= "This region interrupts the coding sequence
FT      of Rep 68 and Rep 40"
FT      857..862
FT      /tag= l
FT      /label= p19_TATA-Box
FT      882..883
FT      /tag= m
FT      /note= "p19 RNA"
FT      1007..2206
FT      /tag= n
FT      /product= "Rep 52"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1007..2272
FT      /tag= o
FT      /product= "Rep 40"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1836..1841
FT      /tag= p
FT      /label= p40_TATA-BOX
FT      1875..1876
FT      /tag= q
FT      /note= "p40 RNA"
FT      2223..4433
FT      /tag= r
FT      /product= "VP1 protein"
FT      /note= "Capsid protein"
FT      2634..4433
FT      /tag= s
FT      /product= "VP2 protein"
FT      /note= "Capsid protein"
FT      /partial
FT      2829..4433
FT      /tag= t
FT      /product= "VP3 protein"
```

```
FT      /note= "Capsid protein"
FT      4447..4452
FT      /tag= u
FT      repeat_unit
FT      4576..4718
FT      /tag= v
FT      /label= 3'-ITR
FT      /note= "Inverted terminal repeat which is capable
FT      of forming T-shaped hairpin structure"
XX      W0200028061-A2.
XX
XX      18-MAY-2000.
XX
XX      02-NOV-1999; 99MO-US25694.
XX
XX      05-NOV-1998; 98US-0107114.
XX      (UYPE-) UNIV PENNSYLVANIA.
XX      Wilson JM, Xiao W;
XX      WPI: 2000-376571/32..
XX      P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
XX      AAY71169.
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host
XX      Claim 1, Fig 1; 108pp; English.
XX
XX      The present sequence is an adeno-associated virus serotype 1 (AAV-1)
XX      DNA characterised by two inverted terminal repeats (ITR) and open
XX      reading frames for rep and capsid (cap) proteins. The rep reading frame
XX      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX      reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX      coding regions, are useful in production of recombinant viral vectors
XX      for gene delivery. These vectors can be used as gene therapy
XX      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX      not induce the formation of neutralising antibodies specific to any
XX      serotype of AAV hence is useful for transforming host cells, and in the
XX      preparation of AAV hence is useful for the delivery of transgene to a host.
XX      Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
XX
XX      Query Match      83.2%; Score 119; DB 21; Length 4718;
XX      Best Local Similarity 89.5%; Pred. No. 2.4e-22;
XX      Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY      1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGTGGGCGCTGCGGACCAAGTCCGC 60
Db      4718 TTGCCCACCTCCCTCTGCGCGCTCGCTCGTGGGCGCGGACGAGAGCTCTGC 4659
QY      61 AGACGGCAGAGCTCTGCTGCGCGGCCACGAGCGAGCGGCGGAGAGGAGTG 120
Db      4658 CGTCTGGGACCTTTGGTCCGAGGCCGCCAGCGAGCGAGCGGCGGAGAGGAGTG 4599
QY      121 GGCACTCCATCACTAGGCTAA 143
Db      4598 GGCACTCCATCACTAGGCTAA 4576
XX
XX      RESULT 5
XX      AAF23748/C
XX      ID AAF23748 standard; DNA; 4722 BP.
XX
XX      AAF23748;
XX      AC
XX      28-MAR-2001 (first entry)
XX      DE AAV3B DNA sequence.
XX      DE
XX      AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
```

KM atherosclerosis; sickle cell anaemia; thalasassaemia;
 KM blood clotting disorder; diabetes; ss.
 OS Adeno associated virus.
 XX US6156303-A.
 XX PD 05-DEC-2000.
 XX PF 11-JUN-1997; 97US-0873168.
 XX PR 11-JUN-1997; 97US-0873168.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Russell DW, Rutledge EA;
 XX DR WPI; 2001-060164/07.
 XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalasassaemia and diabetes
 XX PS Example 2; Fig 1; 50pp; English.
 XX CC The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalasassaemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.
 XX SQ Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;
 XX
 Query Match 79.2%; Score 113.2; DB 22; Length 4722;
 Best Local Similarity 87.3%; Pred. No. 7.8e-21;
 Matches 121; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 1 TTGCCACTCCCTCTGCGGCTCGCTCGCTCGGTGGGGCTTGCGAGCAAAAGTCCGC 60
 Db 4722 TTGGCCACTCCCTCTATGCGCACTCGCTCGGTGGGGCTTGCGAGCAAAAGTCCGC 4663
 QY 61 AGACGGCAGAGCTTGCTTGCGGGCCCGCCACCGAGCGAGCGCGGAGAGAGTG 120
 Db 4662 AGACGGCAGAGCTTGCTTGCGGGCCCGCCACCGAGCGAGCGCGGAGAGTG 4603
 QY 121 GGCACCTCCATCACTAGGGGTA 142
 Db 4602 GCCACTCCATCACTAGAGGTA 4581
 XX
 RESULT 6
 AAF23748
 ID AAF23748 standard; DNA; 4722 BP.
 XX AC AAF23748;
 XX DT 28-MAR-2001 (first entry)
 XX DE AAV3B DNA sequence.
 XX KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 KM atherosclerosis; sickle cell anaemia; thalasassaemia;
 KM blood clotting disorder; diabetes; ss.
 XX OS Adeno associated virus.
 XX US6156303-A.
 XX PD 05-DEC-2000.

XX PF 11-JUN-1997; 97US-0873168.
 XX PR 11-JUN-1997; 97US-0873168.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Russell DW, Rutledge EA;
 XX DR WPI; 2001-060164/07.
 XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalasassaemia and diabetes
 XX PS Example 2; Fig 1; 50pp; English.
 XX CC The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalasassaemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.
 XX SQ Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;
 XX
 Query Match 78.5%; Score 112.2; DB 22; Length 4722;
 Best Local Similarity 87.2%; Pred. No. 1.4e-20;
 Matches 121; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 2 TGCCCACTCCCTCTGCGGCTCGCTCGCTCGGTGGGGCTTGCGAGCAAAAGTCCGC 61
 Db 1 TTGCCACTCCCTCTATGCGCACTCGCTCGGTGGGGCTTGCGAGCAAAAGTCCGC 60
 QY 62 GACGGCAGAGCTTGCTTGCGGGCCCGCCACCGAGCGAGCGCGGAGAGAGTG 121
 Db 61 GACGGCAGAGCTTGCTTGCGGGCCCGCCACCGAGCGAGCGCGGAGAGAGTG 120
 QY 122 GCAACTCCATCACTAGGGGTA 142
 Db 121 CCAACTCCATCACTAGAGGTA 141
 XX
 RESULT 7
 AAQ1448
 ID AAQ1448 standard; DNA; 145 BP.
 XX AC AAQ1448;
 XX DT 27-AUG-1993 (first entry)
 XX DE AAV2 inverted terminal repeat.
 XX KW Adeno-associated virus 2; ITR; site-specific integration; vector;
 KM cell-specific; gene therapy; haemoglobinopathies; thalasassaemia;
 KM diabetes; sickle cell anaemia; cancer; parvovirus; B19; ss.
 XX OS Adeno-associated virus 2.
 XX FH Key Location/Qualifiers
 FT .repeat_region 1..125
 FT /*tag= a
 FT /note= "inverted terminal repeat forming
 FT palindromic hairpin"
 FT misc_feature 42..83
 FT /*tag= b
 FT /note= "Flip orientation"
 XX MO9309239-A.

PA	(RHON) RHONE POULENC RORER SA.
XX	
PI	Denefle P, Latta M, Perricaudet M, Vigne E;
DR	WPI; 1995-320581/41.
XX	
PT	Recombinant defective adenovirus contg. integratable expression cassette - for use in gene therapy to express protein, antigen or anti:sense nucleic acid, also for prodn. of recombinant
PT	adeno-associated viruses
XX	
PS	Example 3; Page 26; 50pp; French.
XX	
CC	Recombinant, non-pathogenic adenovirus which are able to integrate stably into a host genome are claimed. The viruses pref. contain at least one inverted terminal repeat (ITR) sequence and in particular two ITRs flank a heterologous DNA insert. The present sequence is that of the strict ITR from adeno associated virus AAV-2 (i.e. the ITR sequence without any deletions or additions). In the construct pIRRF, the beta-galactosidase marker gene is flanked by two strict AAV-2 ITRs. The defective viruses are useful for stably introducing large fragments of heterologous DNA making them suitable for gene therapy.
CC	
CC	
SO	Sequence 145 BP; 21 A; 52 C; 49 G; 23 T; 0 other;
QY	Query Match 75.1%; Score 107.4; DB 16; Length 145;
Db	Best Local Similarity 85.1%; Prod. No. 2.2e-19;
	Matches 120; Conservative 1; Mismatches 21; Indels 0; Gaps 0;
QY	1 TTGGCCCACTCCCTCTGTGCGCGCTCGCTCGTGGGGCTTGCGGACCAAGTCCGC 60
Db	1 TTGGCCCACTCCCTCTGTGCGCGCTCGCTCGTGAAGCCGGGCAACAAGTCCGC 60
QY	61 AGACGGCAGAGCTCTGCTGTGCCGCCGCCAACCAGCGAGCGCGCAGAAGGGAGTG 120
Db	61 CGAGCGCCCGGGCTTTGGCCCGGGCGGCTCAGTAGAGCGAGCGCGCAGAGAGGAGTG 120
QY	121 GGCAACTTCATCACTAGGGGT 141
Db	121 GCCAACTTCATCACTAGGGGT 141
<hr/>	
RESULT 9	
ID	AAX34295
AC	AAX34295 standard; DNA; 145 BP.
XX	
AC	AAX34295;
DT	
DE	16-JUL-1999 (first entry)
XX	
DE	Adeno-associated virus inverted terminal repeat sequence.
XX	
KM	Recombinant; chimeric; parvovirus; adeno-associated virus; AAV; vector;
KM	promoter; rep; cap; inverted terminal repeat; ITR; erythroid cell;
KM	integration; gene expression; bone marrow; peripheral blood cell;
KM	endothelial cell; myocardial cell; ss.
XX	
OS	Adeno associated virus.
PN	WO9918227-A1.
XX	
PD	15-APR-1999.
XX	
PF	08-OCT-1998; 98WO-US21202
XX	
PR	08-OCT-1997; 97US-0061364
XX	
PA	(ADRE-) ADVANCED RES & TECHNOLOGY INST.
XX	Pomazhagan S, Srivastava A;
XX	WPI; 1999-264033/22.
XX	

XX New recombinant adeno-associated vectors
XX Claim 2, Page 69; 76pp; English.
XX The invention relates to new recombinant chimeric parvovirus-adeno-
XX associated virus (AAV) vectors comprising a promoter e.g. an AAV
XX promoter (AAV34296) and a selected DNA sequence, especially an AAV rep
XX gene and a parvovirus B19 cap gene, located between 2 AAV inverted
XX terminal repeats (ITR) such as the ITR sequence shown here. The system
XX can specifically target primitive progenitor and differentiated cells of
XX the erythroid lineage, and can achieve stable integration and expression
XX of transduced genes. The vectors can be used for the in vitro or in vivo
XX delivery of genes to cells such as bone marrow cells, peripheral blood
XX cells, endothelial cells and myocardial cells.
SQ Sequence 145 BP; 21 A; 52 C; 49 G; 23 T; 0 other;
Query Match 75.1%; Score 107.4; DB 20; Length 145;
Best Local Similarity 85.1%; Pred. No. 2.2e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGACCAAGTCCGC 60
DB 1 TTGCCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGAGGCGCGGACCAAGTCCGC 60
QY 61 AGACGGGAGAGCTCTGCTGCGCGCCGCCACCGAGCGAGCGGCGGAGGAGGAGTG 120
DB 61 CGACGCCCGGCGCTTTGGCCCGCGCGCTCTGAGTGAGCGAGCGCGGAGGAGGAGTG 120
QY 121 GGCACCTCCATCACTAGGGGT 141
DB 121 GCCACCTCCATCACTAGGGGT 141
RESULT 10
AA06769
ID AA06769 standard; DNA; 165 BP.
XX
XX AA06769;
XX 20-JAN-1995 (first entry)
XX Double-D ITR terminal repeat.
XX
XX Double-D ITR; inverted terminal repeat; D-sequence; AAV;
XX adeno-associated virus; capsid; encapsidation; gene therapy; vector;
XX de.
XX
XX Synthetic.
XX
XX WO9413788-A.
XX
XX 23-JUN-1994.
XX
XX 03-DEC-1993; 93WO-US11728.
XX
XX 04-DEC-1992; 92US-0989841.
XX
XX (UYP1-) UNIV PITTSBURGH.
XX
XX Samulski RJ, Xiao X;
XX WPI; 1994-217868/26.
XX
XX Adenovirus associated viral inverted terminal repeat - for use in
XX a recombinant viral vector system for treatment of genetic
XX diseases
XX
XX Disclosure; Page 25; 44pp; English.
XX The 20-bp D-sequence given in AA06773 is present in the inverted
XX terminal repeat (ITR) sequence of AAV and is required for viral

CC replication. A novel, modified terminal repeat structure, double-D
CC ITR, was constructed (AA06769) that contained a single 145 bp ITR
CC sequence with an additional D' sequence. The double-D ITR fragment
CC allows replication and encapsidation of recombinant DNA into AAV
CC capsid proteins. Vectors including double-D ITR can be used for
CC gene replacement therapies.
XX
SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;
Query Match 75.1%; Score 107.4; DB 15; Length 165;
Best Local Similarity 85.1%; Pred. No. 2.2e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCCACTCCCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGACCAAGTCCGC 60
DB 21 TTGCCCCACTCCCTCTGCGCGCTCGCTCGCTCGGTGAGGCGCGGACCAAGTCCGC 80
QY 61 AGACGGGAGAGCTCTGCTGCGCGCCGCCACCGAGCGAGCGGCGGAGGAGGAGTG 120
DB 81 CGACGCCCGGCGCTTTGGCCCGCGCGCTCTGAGTGAGCGAGCGCGGAGGAGGAGTG 140
QY 121 GGCACCTCCATCACTAGGGGT 141
DB 141 GCCACCTCCATCACTAGGGGT 161
RESULT 11
AAT9462
ID AAT9462 standard; DNA; 165 BP.
XX
XX AAT9462;
XX 10-SEP-1997 (first entry)
XX Adeno associated virus inverted terminal repeat with double D region.
XX
XX Viral replication; RBP protein; inverted terminal repeat; ITR;
XX adeno-associated virus; AAV; vector; double-D; cis-acting;
XX lytic life cycle; gene therapy; ss.
XX
XX Adeno associated virus.
XX
XX Synthetic.
XX
XX Key
XX
XX repeat_unit
XX
XX 1..145
XX /rpt_type= INVERTED
XX /standard_name= ITR
XX /note= "This naturally occurring 145 bp ITR is
XX located at both ends of the AAV genome"
XX
XX 1..20
XX /rpt_type= INVERTED
XX /label= D_repeat
XX 21..62
XX /rpt_type= INVERTED
XX /label= A_repeat
XX /note= "Forms the stem of a T-shaped structure when
XX base paired with repeat A"
XX
XX repeat_unit
XX
XX 63..70
XX /rpt_type= INVERTED
XX /label= C_repeat
XX 74..81
XX /rpt_type= INVERTED
XX /label= C_repeat
XX 85..92
XX /rpt_type= INVERTED
XX /label= B_repeat
XX 96..103
XX repeat_unit

```
FT      /*tag= 9
FT      /rpt_type= INVERTED
FT      /label= B' _repeat
FT      104..145
FT      repeat_unit
FT      /*tag= h
FT      /rpt_type= INVERTED
FT      /label= A' repeat
FT      /note= "Forms the stem of a T-shaped structure when
FT      base paired with A' repeat"
FT      repeat_unit
FT      146..165
FT      /*tag= 1
FT      /label= D'
FT      /note= "Additional D' sequence"
FT      61..82
FT      /*tag= 1
FT      /note= "Part of T-shaped structure, formed by base
FT      pairing between the C' and C repeats"
FT      stem_loop
FT      84..104
FT      /*tag= k
FT      /note= "Part of T-shaped structure, formed by base
FT      pairing between the B' and B repeats"
FT      misc_structure
FT      21..145
FT      /*tag= 1
FT      /label=
FT      /note= "The naturally occurring 145 bp ITR sequence
FT      can form a T-shaped structure for DNA
FT      replication when single-stranded"
FT      repeat_region
FT      1..165
FT      /*tag= m
FT      /label= Double-D
FT      /note= "This 165 bp sequence, resulting from
FT      addition of a second D repeat to the AAV
FT      ITR, has not been identified in any
FT      naturally occurring virus"
XX      MO9636364-A1.
XX      21-NOV-1996.
XX      14-MAY-1996; 96MO-US06786.
XX      15-MAY-1995; 95US-0440738.
XX      (SAMU/) SAMULSKI R J.
XX      (XIAO/) XIAO X.
XX      PA
XX      PI Samulski RJ, Xiao X;
XX      DR WPI; 1997-042643/04.
XX      Double-D sequence directs adeno-associated virus integration into
XX      host genome - used in gene therapy, maintains full length coding
XX      sequence of therapeutic gene
XX      PS
XX      Claim 1; Fig 9; 55pp; English.
XX      The adeno associated virus (AAV) has a 145 bp inverted terminal
XX      repeat (ITR) located at each end of its genome. In addition to
XX      being able to base pair with each other, the ITRs can also
XX      individually fold back on themselves through the base pairing of A,
XX      A', B, B' and C, C' sequences to form a T-shaped structure for DNA
XX      replication (see features table). It has been found that viral
XX      mutants with deleted D sequences are unable to replicate their DNA.
XX      Addition of a second D repeat (D') to the naturally occurring ITR
XX      resulted in a sequence, designated double-D, which was sufficient
XX      to carry out the functions normally required of two wild-type ITRs
XX      during a lytic AAV viral infection, i.e. it is capable of directing
XX      replication and assembly into AAV, and/or the integration into the
XX      host genome, of recombinant DNA containing the nucleic acid molecule.
XX      Vectors and viral particles containing the double-D sequence are
XX      useful in gene therapy. Replication and integration into the host
XX      genome is completely effected through the double-D sequences, ensuring
XX      that the heterologous gene sequences remain intact.
```

```
XX      SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;
XX      Query Match 75.1%; Score 107.4; DB 18; Length 165;
XX      Best Local Similarity 85.1%; Pred. No. 2.2e-19;
XX      Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY      1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCGCTGAGGCGCTGCGGACCAAGTCCG 60
DB      21 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGCTCGCTGAGGCGCGGACCAAGTCCG 80
QY      61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCGCCACGACGCGCGGACGAGGAGTG 120
DB      81 CGAGCGCCGGGCTTGGCCCGGCGGCGCTCAGTGAGCGAGCGAGCGGACGAGGAGTG 140
QY      121 GCGAATCTCCATCACTAGGGGT 141
DB      141 GCGAATCTCCATCACTAGGGGT 161
RESULT 12
AD03535
ID AD03535 standard; DNA; 165 BP.
AC AD03535;
XX      19-JUN-2001 (first entry)
DT
XX      Inverted terminal repeat double DD DNA sequence.
DE
XX      Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;
XX      inverted terminal repeat; haemostatic; antiscikling; neuroprotective;
XX      antianaemic; noctropic; blood disorder; sickle cell anaemia;
XX      thalassemia; neurological disorder; haemophilia; Alzheimer's disease;
XX      muscle disorder; Parkinson's disease; gene delivery; erythropoietin;
XX      epo; CFT; cystic fibrosis transmembrane conductance receptor;
XX      tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;
XX      gene therapy; Double DD; ds.
XX      OS
XX      Unidentified.
XX      MO200125465-A1.
XX      12-APR-2001.
XX      06-OCT-2000; 2000MO-US27863.
XX      07-OCT-1999; 99US-0158209.
XX      (IOWA ) UNIV IOWA RES FOUND.
XX      (ENGEL/) ENGELHARDT J F.
XX      (DONG/) DONGSHENG D.
XX      (ZIYI/) ZIYING Y.
XX      PI Engelhardt JF, Dongsheng D, Ziyang Y;
XX      DR WPI; 2001-266321/27.
XX      Composition for transferring recombinant DNAs and to express a
XX      polypeptide in a host cell, comprises two recombinant adeno-associated
XX      viruses -
XX      Disclosure; Page 139; 144pp; English.
XX      The patent discloses a composition comprising at least two recombinant
XX      adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA
XX      molecule comprising three DNA segments linked together. The first and
XX      the third DNA segments comprise 5' and 3' inverted terminal repeats
XX      (ITRs) respectively from the circular intermediate of AAV. The second
XX      DNA segment in each virus is different and does not comprise AAV
XX      sequence. It preferably comprises sequences encoding a therapeutically
XX      effective polypeptide such as the cystic fibrosis transmembrane
XX      conductance receptor gene (CFTR), the erythropoietin (epo) gene, the
```

CC Tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV
CC imparts increased episomal stability and persistence of the vector in
CC the host cell. Compositions comprising rAAV sequences are useful for
CC transferring recombinant DNAs to a host cell and express a polypeptide
CC in a host cell. The recombinant vector is useful in medical therapy,
CC which includes treatment or prophylaxis of blood disorders (e.g. sickle
CC cell anaemia, thalassemia, haemophilia), neurological disorders, such
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery
CC vehicle for gene therapy.
CC The present sequence is the ITR DD DNA sequence referred to as
CC "double sequence". This sequence is disclosed in U.S. Patent No.
CC 5,478,745. This sequence is not used in the present invention.

SO Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 165;
Best Local Similarity 85.1%; Pred. No. 2.2e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGCGAACCAAGTCCGC 60
DB 21 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTAGGCGCGCAACCAAGTCCGC 80
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCCAACGAGCGAGCGCGCAGAGGAGTG 120
DB 81 CGACGCCCGGGCTTTGGTGGCGCGCTCACTAGAGCAGAGCGCGCAGAGGAGTG 140
QY 121 GGCACCTCCATCACTAGGGGT 141
DB 141 GCCAAGCTCCATCACTAGGGGT 161

RESULT 13
AAZ47166
ID AAZ47166 standard; DNA; 272 BP.

XX AAZ47166;

DT 28-MAR-2000 (first entry)

DE Head-to-tail ITR sequence from recombinant AAV vector p1202.

XX Antisickling; hemostatic; neuroprotective; antiparkinsonian; ITR;
KM gene therapy; inverted terminal repeat; vector; episomal stability;
KM blood; neurological; muscle; disorder; ds.

XX Adeno associated virus.

OS

XX WO960146-A1.

XX 25-NOV-1999.

XX 20-MAY-1999; 99WO-US11197.

XX 20-MAY-1998; 98US-0086166.

XX 25-MAR-1999; 99US-0276625.

XX (IOWA) UNIV IOWA RES FOUND.

XX (ENGE/) ENGELHARDT J F.

XX (DUAN/) DUAN D.

XX Engelhardt JF, Duan D;

XX WPI; 2000-062462/05.

XX Recombinant adeno-associated virus vector useful for gene therapy

PT against disorders related to blood, neurological and muscular systems

XX Disclosure; Fig 10C; 121pp; English.

CC This sequence corresponds to a head-to-tail inverted terminal repeat
CC (ITR) sequence from the recombinant adeno-associated virus (rAAV)
CC vector p1202. The invention relates to a DNA molecule comprising a DNA
CC segment (or biologically active subunit or variant) of a circular
CC intermediate of an AAV that confers increased episomal stability,
CC #ersistence or abundance in a host cell. Compositions containing a
CC vector with therapeutic gene and delivery vehicle or containing two
CC vectors expressing a full length polypeptide coordinately, are useful
CC for manufacturing a medication for treating pathological conditions or
CC symptoms in a mammal. The vector is useful for therapeutic or
CC prophylactic treatments of blood disorders (e.g. sickle cell anemia,
CC thalassemia, hemophilias and Fanconi's anemia), neurological disorders
CC (e.g Alzheimer's disease, Parkinson's disease) and muscle disorders.

SO Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;

Query Match 75.1%; Score 107.4; DB 21; Length 272;
Best Local Similarity 85.1%; Pred. No. 2.3e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGCGAACCAAGTCCGC 60
DB 69 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTAGGCGCGCAACCAAGTCCGC 128
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCCAACGAGCGAGCGCGCAGAGGAGTG 120
DB 129 CGACGCCCGGGCTTTGGTGGCGCGCTCACTAGAGCAGAGCGCGCAGAGGAGTG 168
QY 121 GGCACCTCCATCACTAGGGGT 141
DB 189 GCCAAGCTCCATCACTAGGGGT 209

RESULT 14
AAD03534
ID AAD03534 standard; DNA; 272 BP.

XX AAD03534;

DT 19-JUN-2001 (first entry)

DE AAV circular intermediate DNA segment comprising ITRs from clone p1202.

XX Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;
KM inverted terminal repeat; haemostatic; antisickling; neuroprotective;
KM antianaemic; noctropic; blood disorder; sickle cell anaemia;
KM thalassemia; neurological disorder; haemophilia; Alzheimer's disease;
KM muscle disorder; Parkinson's disease; gene delivery; erythropoietin;
KM epo; CFTR; cystic fibrosis transmembrane conductance receptor;
KM tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;
KM gene therapy; clone p1202; ds.

XX

XX Adeno associated virus.

OS

XX WO200125465-A1.

XX 12-APR-2001.

XX 06-OCT-2000; 2000WO-US27863.

XX 07-OCT-1999; 99US-0158209.

XX (IOWA) UNIV IOWA RES FOUND.

XX (ENGE/) ENGELHARDT J F.

XX (DONG/) DONGSHENG D.

XX (ZIYI/) ZIYING Y.

XX Engelhardt JF, Dongsheng D, Ziyang Y;

XX WPI; 2001-266321/27.

XX Composition for transferring recombinant DNAs and to express a

PT polypeptide in a host cell, comprises two recombinant adeno-associated

PT viruses -
XX
PS Disclosure: Fig 10C, 144pp; English.
XX
CC The patent discloses a composition comprising at least two recombinant
CC adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA
CC molecule comprising three DNA segments linked together. The first and
CC the third DNA segments comprise 5' and 3' inverted terminal repeats
CC (ITRs) respectively from the circular intermediate of AAV. The second
CC DNA segment in each virus is different and does not comprise AAV
CC sequence. It preferably comprises sequences encoding a therapeutically
CC effective polypeptide such as the cystic fibrosis transmembrane
CC conductance receptor gene (CFTR), the erythropoietin (epo) gene, the
CC tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV
CC imparts increased episomal stability and persistence of the vector in
CC the host cell. Compositions comprising rAAV sequences are useful for
CC transferring recombinant DNAs to a host cell and express a polypeptide
CC in a host cell. The recombinant vector is useful in medical therapy,
CC which includes treatment or prophylaxis of blood disorders (e.g. sickle
CC cell anaemia, thalassemia, haemophilia), neurological disorders, such
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery
CC vehicle for gene therapy.
CC The present sequence is circular intermediate DNA segment comprising
CC ITRs from Adeno associated virus (AAV) clone p1202.
XX
SQ Sequence 272 BP, 58 A, 78 C, 78 G, 58 T, 0 other;
Query Match 75.1%; Score 107.4; DB 22; Length 272;
Best Local Similarity 85.1%; Pred. No. 2.3e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCCACTCCCTCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 128
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 120
DB 129 CGAGCGCCCGGCGCTTTGGTGGCGCCGCTCAGTGAAGCGAGCGCGGAGAGGAGTG 188
QY 121 GGCAACTCCATCTAGGGGT 141
DB 189 GCCAACTCCATCTAGGGGT 209
RESULT 15
AAD37254
ID AAD37254 standard; DNA; 955 BP.
XX
AC AAD37254;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid #1.
XX
KM Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KM adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KM Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.

PI XIAO X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 57; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid containing human dystrophin minigene, a muscle creatine
CC kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 955 BP, 177 A; 307 C; 296 G; 175 T; 0 other;
Query Match 75.1%; Score 107.4; DB 24; Length 955;
Best Local Similarity 85.1%; Pred. No. 2.4e-19;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCCACTCCCTCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 60
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCTGCTCGCTCGGTGGGGGCTTGGGACCAAGTCCGC 120
DB 61 CGAGCGCCCGGCGCTTTGGCGCGGCGCTCAGTGAAGCGAGCGCGGAGAGGAGTG 120
QY 121 GGCAACTCCATCTAGGGGT 141
DB 121 GCCAACTCCATCTAGGGGT 141
Search completed: July 20, 2003, 09:39:47
Job time : 32.1207 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 08:54:01 ; Search time 6.68705 Seconds

(without alignments)
6558.165 Million cell updates/sec

Title: US-09-807-802a-1_COPY_1_143

Perfect score: 143

Sequence: 1 ttggccacacccctctctgcgc.....aactccatcactaggggtaaa 143

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.4	75.1	145	1	US-07-789-917A-1
2	107.4	75.1	145	3	US-08-702-573-4
3	107.4	75.1	145	4	US-07-982-193-1
4	107.4	75.1	165	1	US-07-989-841A-1
5	107.4	75.1	165	2	US-08-440-738A-1
6	107.4	75.1	165	3	US-08-471-914-1
7	107.4	75.1	165	4	US-09-276-625-7
8	107.4	75.1	272	4	US-09-276-625-6
9	107.4	75.1	4680	1	US-08-254-358-1
10	107.4	75.1	4680	1	US-08-475-391-1
11	107.4	75.1	4680	2	US-08-709-609-1
12	107.4	75.1	4680	5	PCT-US95-07178-1
13	107.4	75.1	5932	4	US-09-299-141-4
14	107.4	75.1	5932	4	US-09-299-141-4
15	107.4	75.1	6142	4	US-09-299-141-8
16	107.4	75.1	6142	4	US-09-299-141-8
17	107.4	75.1	6253	3	US-08-893-327-15
18	107.4	75.1	6253	3	US-08-893-327-15
19	107.4	75.1	6280	3	US-08-893-327-17
20	107.4	75.1	6280	3	US-08-893-327-17
21	107.4	75.1	6280	3	US-08-893-327-19
22	107.4	75.1	6280	3	US-08-893-327-19
23	107.4	75.1	6565	4	US-09-299-141-1
24	107.4	75.1	6565	4	US-09-299-141-1
25	107.4	75.1	6714	4	US-09-299-141-6
26	107.4	75.1	6714	4	US-09-299-141-6
27	107.4	75.1	6924	4	US-09-299-141-9

C	28	107.4	75.1	6924	4	US-09-299-141-9	Sequence 9, Appli
C	29	107.4	75.1	6924	4	US-09-299-141-10	Sequence 10, Appl
C	30	107.4	75.1	6924	4	US-09-299-141-10	Sequence 10, Appl
C	31	107.4	75.1	6924	4	US-09-299-141-11	Sequence 11, Appl
C	32	107.4	75.1	6924	4	US-09-299-141-11	Sequence 11, Appl
C	33	107.4	75.1	6981	4	US-09-299-141-7	Sequence 7, Appli
C	34	107.4	75.1	7054	4	US-09-299-141-3	Sequence 3, Appli
C	35	107.4	75.1	7054	4	US-09-299-141-3	Sequence 3, Appli
C	36	107.4	75.1	7405	4	US-09-299-141-2	Sequence 2, Appli
C	37	107.4	75.1	7405	4	US-09-299-141-2	Sequence 2, Appli
C	38	107.4	75.1	7492	4	US-09-299-141-5	Sequence 5, Appli
C	39	107.4	75.1	7492	4	US-09-299-141-5	Sequence 5, Appli
C	40	107.4	75.1	8638	4	US-09-276-625-4	Sequence 4, Appli
C	41	107.4	75.1	8638	4	US-09-276-625-4	Sequence 4, Appli
C	42	105.8	74.0	272	4	US-09-276-625-6	Sequence 6, Appli
C	43	104.2	72.9	5585	2	US-08-305-221-1	Sequence 1, Appli
C	44	102.6	71.7	272	4	US-09-276-625-6	Sequence 6, Appli
C	45	99.4	69.5	145	4	US-08-525-866-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-07-789-917A-1
Sequence 1, Application US/0789917A
Patent No. 5252479
GENERAL INFORMATION:
APPLICANT: Srivastava, Arun
TITLE OF INVENTION: SAFE VECTOR FOR GENE THERAPY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release q.0, Version q.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789, 917A
FILING DATE: 19911118
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McNulty, William B.
REGISTRATION NUMBER: 22,606
REFERENCE/DOCKET NUMBER: 8361
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 base pairs
TYPE: NUCLEIC ACID
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-789-917A-1
Query Match 75.1%; Score 107.4; DB 1; Length 145;
Best Local Similarity 85.1%; Pred. No. 2, 1e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGGCCACTCTCTCTGCGCGCTGCGTCTGCGTGGGCGCTCGGACCAAGGTCGCC 60
Db 1 TTGGCCACTCTCTCTGCGCGCTGCGTCTGCGTCTGCGTGGGCGCGGCGCAAGGTCGCC 60
QY 61 AGACGGAGAGCTCTGCTCTGCGGCGCCCGACGAGCGAGCGGCGGCAAGAGGATG 120

Db 61 CGAGCGCCGGGGCTTTTGCCCGGGCGGCTCATGTAGCGAGCGCA3CCGCGAGAGAGGGAATG 120
 QY 121 GGCACCTCCATCACTAGGGGT 141
 Db 121 GGCACCTCCATCACTAGGGGT 141

RESULT 2
THE 09-70

Sequence 4 Application US/08702573
Patent No. 6033885
GENERAL INFORMATION:
APPLICANT: LATTA, Martine
APPLICANT: DENEFLÉ, Patrice
APPLICANT: VIGNE, Emmanuelle
APPLICANT: PERRICAUDET, Michel
TITLE OF INVENTION: INTEGRATIVE RECOMBINANT ADENOVIRUSES,
TITLE OF INVENTION: PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Rd. 3c43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08702.573
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94/02445
FILING DATE: 03-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/00233
FILING DATE: 28-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Smith Ph.D., Julie K.
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: S794011-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..145
OTHER INFORMATION: /note= "Minimal TTR Sequence"
US-08-702-573-4

[illegible]

QY	121	GGCACTCCATCACTAGGGT	141
Db	121	GCCAAGTCATCACTAGGGT	141

RESULT 3

Sequence 1, Application US/07982193
Patent No. 6261834
GENERAL INFORMATION:
APPLICANT: Sivasava, Arun
TITLE OF INVENTION: SAFE VECTOR FOR GENE THERAPY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/07/982,193
FILING DATE: 19921125
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McNulty, William E.
REGISTRATION NUMBER: 22,606
REFERENCE/DOCKET NUMBER: 8361
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-982-193-1

Query Match	75.1%;	Score 107.4;	DB 4;	Length 145;
Best Local Similarity	85.1%;	Pred. No. 2.1e-21;		
Matches	120;	Conservative	0;	Mismatches 21; Indels 0; Gaps 0;
Qy	1	TTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTCGGGCGTCGCGAGCAACCAAGTCCGC	60	
Db	1	TTGGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTAGAGCGCGGCGACCAAGTCCGC	60	
Qy	61	AGACGCGAAGACTCTGCTCTGCGCGGCCCCACCGACGACGAGCGCGCGACAGAGAGGAGTG	120	
Db	61	CGACGCGCCGGGCTTTGCCCGGCGGCTCACTGAGCGACGAGCGCGCAGAGAGGAGTG	120	
Qy	121	GGCAACTTCATCACTAGGGGT	141	
Db	121	GCCAACTTCATCACTAGGGGT	141	

RESULT 4

Sequence 1: Application US/07968984A
Patent No. 5476745
GENERAL INFORMATION:
APPLICANT: Samulski, R. J.
APPLICANT: Xiao, X.
TITLE OF INVENTION: Recombinant Viral Vector System
NUMBER OF SEQUENCES: 6

APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,391
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/254,358
ATTORNEY/AGENT INFORMATION:
NAME: No. 578621land, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-475-391-1

Query Match 75.1%; Score 107.4; DB 1; Length 4680;
Best Local Similarity 85.1%; Pred. No. 3e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGCAAAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTGTGCGGCGCCACCGAGCGAGCGGCGCGAGAGGAGTG 120
DB 61 CGACGCCCGGGCTTTGCGCGCGGCGCTCACTGAGCGAGCGCGCGAGAGGAGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GCCAATCCATCACTAGGGGT 141

RESULT 11
US-08-709-609-1
Sequence 1, Application US/08709609
Patent No. 5858775
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709,609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5858775and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-709-609-1

Query Match 75.1%; Score 107.4; DB 2; Length 4680;
Best Local Similarity 85.1%; Pred. No. 3e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGCAAAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTGTGCGGCGCCACCGAGCGAGCGGCGCGAGAGGAGTG 120
DB 61 CGACGCCCGGGCTTTGCGCGCGGCGCTCACTGAGCGAGCGCGCGAGAGGAGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GCCAATCCATCACTAGGGGT 141

RESULT 12
PCT-US95-07178-1
Sequence 1, Application PC/TUS9507178
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07178
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300

```

?
? TELEX: (312) 474-0448
? TEL#: 25-3856
?
? INFORMATION FOR SEQ ID NO: 1:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 4680 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
PCT-US95-07176-1

```

Query Match	75.1%;	Score 107.4;	DB 5;	Length 4680;
Best Local Similarity	85.1%;	Pred. No. 3e-21;		
Matches 120; conservative	0;	Mismatches 21;	Indels 0;	Gaps 0

Oy	TTGGCCACTCCCTCTGTGGCGCGCTGCCTGCTCGGTTGGGGGCTTGGGAGCAAAAGTCCGC	60
Db	1 TTGGCCACTCCCTCTGTGGCGCGCTGCCTGCTACTGAGGCGCGCGCAAAAGTCCGC	60
Oy	61 AGAGCGAGAGACTCTGCTGTGCGGGGCCACCGAGCGAGCGAGCGAGAGAGGAGTG	120
Db	61 CGAGCGCCGGGCTTTGGCCCGGGCGGCTCTAGTGAAGCAAGCGCGCGAGAGAGGAGTG	120
Oy	121 GGCAACTCCACTACTAGGGGT	141
Db	121 GCCAACTCCACTACTAGGGGT	141

RESULT 13
US-09-299-141-4

```

Patent No. 5461506
GENERAL INFORMATION:
APPLICANT: FLOTE, TERENCE R.
APPLICANT: SONG, STRONG
APPLICANT: BYRNE, BARRY J.
APPLICANT: MORGAN, MICHAEL
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY
FILE REFERENCE: 4300.011800
CURRENT APPLICATION NUMBER: US/09/299,141
CURRENT FILING DATE: 1999-04-23
EARLIER APPLICATION NUMBER: 60/083,025
EARLIER FILING DATE: 1998-04-24
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn, Ver. 2.0
SEQ ID NO 4
LENGTH: 5932
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:p43
US-09-299-141-4

```

Query Match	75.1%	Score 107.4	DB 4	length 5932
Best Local Similarity	85.1%	Pred. No. 3.1e-21		
Matches 120; Conservative	0	Mismatches 21	indels 0	Gaps 0

Oy	1	TTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGCGGAGCCATGGGAACTCAAAGGTCCG	60
	18	TTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGCTGCGGAGCCAAAGGTCCG	77
Db	1	TTGGCCACTCCCTCTGCGCGCTGCTGCTGCTGCTGCGGAGCCAAAGGTCCG	77
Oy	61	AGACGCGAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGCGGAGAGAGGAGTG	120
	78	CGAGCGCCCGGAGCTTTGGCCCGGCGGCGCTCTAGTGAAGCGAGCGCGGAGAGGAGTG	137
Db	1	AGACGCGAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGCGGAGAGAGGAGTG	120
Oy	121	GGCAACTCATCACTAGGGGT	141
	138	GCCAACTCACTAGGGGT	158
Db	1	GGCAACTCATCACTAGGGGT	141
	138	GCCAACTCACTAGGGGT	158

RESULT 14
US-09-299-141-4/C
; Sequence 4, Application US/09299141

Patent No. 6461606
GENERAL INFORMATION:
APPLICANT: FLOTTE, TERENCE R.
APPLICANT: SONG, SIKHONG
APPLICANT: BYRNE, BARRY J.
APPLICANT: MORGAN, MICHAEL
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY
FILE REFERENCE: 4300, 011800
CURRENT APPLICATION NUMBER: US/09/299,141
CURRENT FILING DATE: 1999-04-23
EARLIER APPLICATION NUMBER: 60/083,025
EARLIER FILING DATE: 1998-04-24
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 5932
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: p43C-AT
US-09-299-141-4

Query Match	75.1%;	Score 107.4;	DB 4	Length 5932;
Best Local Similarity	85.1%;	Pred. No. 3.1e-21;		
Matches 120;	Conservative	0;	Mismatches 21;	Indels 0;
				Gaps 0;

OY	1	TTGGCCCACTCCCTCTCTGCGGCGTCTGCTGGTGGTGGGGGCTGGCGACCCAAAGGTCCG	60
Db	3078	TTGGCCCACTCCCTCTCTGCGGCGTCTGCTGGTGGTGGGGGCTGGCGACCCAAAGGTCCG	3015
OY	61	AGAGGGGAGAGCTGCTGCTGCGGCGCCCAACGAGCGGAGCGCGGAGAGAGGGAGTG	120
Db	3018	CGAGCGCCCGGGCTTTGGCCCGGGGCGGCTCACTGAGCGAGCGGCGCGGAGAGGGAGTG	2955
OY	121	GGCAACTCCATCACTAGGGGT	141
Db	2958	GCCAACTCCATCACTAGGGGT	2938

RESULT 15
US-09-299-141-8
: Sequence 8, Application US/09299141

```

GENERAL INFORMATION:
PATENT NO.: 0901000
APPLICANT: FLOTTE, TERENCE R.
APPLICANT: SONG, SIHONG
APPLICANT: BYRNE, BARRY J.
APPLICANT: MORGAN, MICHAEL
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY
FILE REFERENCE: 4300, 011800
CURRENT APPLICATION NUMBER: US/09/299,141
CURRENT FILING DATE: 1999-04-23
EARLIER APPLICATION NUMBER: 60/083,025
EARLIER FILING DATE: 1996-04-24
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 6142
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: PLASMID
US-09-299-141-8

```

Query Match	75.1%	Score 107.4	DB 4	Length 6142
Best Local Similarity	85.1%	Pred. No. 3.1e-21		
Matches 120	Conservative	0	Mismatches 21	Indels 0
				Gaps 0

Dy		1 TTGCCACTCCCTCTGCGCGCTGGTGGTGAGCAAAAGGTTCGC	60
Ddb		18 TTGGCACA TCCCTCTGCGCGCTGGTGGTGAAGGCGA CCAAAGGTTCGC	77

Qy	61	AGAGGCGAGAGCTCTGCTGCGGCCCA	CGAGCGAGCGCGCGAGAGGAGTG	120
Db	78	CGACGCGCGGCTTTGCGCGCGGCTCA	GTGAGCGAGCGCGCGAGAGGAGTG	137
Qy	121	GGCACTTCATCACTAGGGGT		141
Db	138	GCCAACTTCATCACTAGGGGT		158

Search completed: July 20, 2003, 13:23:48
 Job time : 6.68705 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 13:19:56 ; Search time 30.7204 Seconds
(without alignment)

9603.029 million cell updates/sec

Title: US-09'-807-802A-1_COPY_1_143

Sequence: 1 ttgccactccctctctgcg.....aactccatcactaggggtaa 143

Scoring table: IDENTITY_NUC

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/prodata/1/pubmap/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubmap/US07_PUBCOMB.seq:*
- 3: /cgn2_6/prodata/1/pubmap/PCR_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubmap/US06_NEW_PUB.seq:*
- 5: /cgn2_6/prodata/1/pubmap/US06_PUBCOMB.seq:*
- 6: /cgn2_6/prodata/1/pubmap/US07_NEW_PUB.seq:*
- 7: /cgn2_6/prodata/1/pubmap/PCRUS_PUBCOMB.seq:*
- 8: /cgn2_6/prodata/1/pubmap/US08_NEW_PUB.seq:*
- 9: /cgn2_6/prodata/1/pubmap/US08_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubmap/US09_NEW_PUB.seq:1
- 11: /cgn2_6/prodata/1/pubmap/US09_NEW_PUB.seq:2
- 12: /cgn2_6/prodata/1/pubmap/US09_NEW_PUB.seq:3
- 13: /cgn2_6/prodata/1/pubmap/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubmap/US10_NEW_PUB.seq:*
- 15: /cgn2_6/prodata/1/pubmap/US10_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubmap/US60_NEW_PUB.seq:*
- 17: /cgn2_6/prodata/1/pubmap/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	107.4	75.1	145	11	US-09-782-378A-6	Sequence 6, Appl
2	107.4	75.1	146	15	US-10-135-984-8	Sequence 8, Appl
3	107.4	75.1	165	11	US-09-782-378A-8	Sequence 8, Appl
4	107.4	75.1	165	15	US-10-054-665-7	Sequence 7, Appl
5	107.4	75.1	207	15	US-10-023-208-58	Sequence 58, Appl
6	107.4	75.1	272	15	US-10-054-665-6	Sequence 6, Appl
7	107.4	75.1	4675	11	US-09-782-378A-1	Sequence 1, Appl
8	107.4	75.1	4675	11	US-09-782-378A-2	Sequence 2, Appl
9	107.4	75.1	4675	15	US-10-240-196-1	Sequence 1, Appl
10	107.4	75.1	4679	10	US-09-804-898-1	Sequence 1, Appl
11	107.4	75.1	4679	10	US-09-945-681-10	Sequence 10, Appl
12	107.4	75.1	4679	15	US-10-038-972A-12	Sequence 12, Appl
13	107.4	75.1	4680	15	US-10-077-294-1	Sequence 1, Appl
14	107.4	75.1	4680	15	US-10-163-886-1	Sequence 1, Appl
15	107.4	75.1	4680	15	US-10-263-127-1	Sequence 1, Appl
16	107.4	75.1	5932	15	US-10-267-117-4	Sequence 4, Appl

C 17	107.4	75.1	5932	15	US-10-267-117-4	Sequence 4, Appl1
C 18	107.4	75.1	5932	15	US-10-340-112-4	Sequence 4, Appl1
C 19	107.4	75.1	5932	15	US-10-340-112-4	Sequence 4, Appl1
C 20	107.4	75.1	6142	15	US-10-267-117-8	Sequence 8, Appl1
C 21	107.4	75.1	6142	15	US-10-267-117-8	Sequence 8, Appl1
C 22	107.4	75.1	6142	15	US-10-340-112-8	Sequence 8, Appl1
C 23	107.4	75.1	6142	15	US-10-340-112-8	Sequence 8, Appl1
C 24	107.4	75.1	6565	15	US-10-267-117-1	Sequence 1, Appl1
C 25	107.4	75.1	6565	15	US-10-267-117-1	Sequence 1, Appl1
C 26	107.4	75.1	6565	15	US-10-340-112-1	Sequence 1, Appl1
C 27	107.4	75.1	6565	15	US-10-340-112-1	Sequence 1, Appl1
C 28	107.4	75.1	6714	15	US-10-267-117-6	Sequence 6, Appl1
C 29	107.4	75.1	6714	15	US-10-267-117-6	Sequence 6, Appl1
C 30	107.4	75.1	6714	15	US-10-340-112-6	Sequence 6, Appl1
C 31	107.4	75.1	6714	15	US-10-340-112-6	Sequence 6, Appl1
C 32	107.4	75.1	6924	15	US-10-267-117-9	Sequence 9, Appl1
C 33	107.4	75.1	6924	15	US-10-267-117-9	Sequence 9, Appl1
C 34	107.4	75.1	6924	15	US-10-267-117-10	Sequence 10, Appl1
C 35	107.4	75.1	6924	15	US-10-267-117-10	Sequence 10, Appl1
C 36	107.4	75.1	6924	15	US-10-267-117-11	Sequence 11, Appl1
C 37	107.4	75.1	6924	15	US-10-267-117-11	Sequence 11, Appl1
C 38	107.4	75.1	6924	15	US-10-340-112-9	Sequence 9, Appl1
C 39	107.4	75.1	6924	15	US-10-340-112-9	Sequence 9, Appl1
C 40	107.4	75.1	6924	15	US-10-340-112-10	Sequence 10, Appl1
C 41	107.4	75.1	6924	15	US-10-340-112-10	Sequence 10, Appl1
C 42	107.4	75.1	6924	15	US-10-340-112-11	Sequence 11, Appl1
C 43	107.4	75.1	6924	15	US-10-340-112-11	Sequence 11, Appl1
C 44	107.4	75.1	6981	15	US-10-267-117-7	Sequence 7, Appl1
C 45	107.4	75.1	6981	15	US-10-267-117-7	Sequence 7, Appl1

ALIGNMENTS

```

RESULT 1
US-09-782-378A-6
Sequence 6, Application US/09782378A
Patent No. US20020102731A1
GENERAL INFORMATION:
APPLICANT: Hearing, Patrick
APPLICANT: Babou, Madie
APPLICANT: Sandelko, Ziv
APPLICANT: Gatenko, Dmitri
TITLE OF INVENTION: Adenoviral Vectors
FILE REFERENCE: STONB-04970
CURRENT APPLICATION NUMBER: US/09/782,378A
CURRENT FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/227,747
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 145
TYPE: DNA
ORGANISM: Homo sapiens
US-09-782-378A-6

```

Query Match	75.1%;	Score 107.4;	DB 11;	Length 145;
Best Local Similarity	85.1%;	Pred. No. 2.8e-21;		
Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

OY	1	TTGCCCACCTCCCTCTCTGTGGCGCTGGCTCGCTCGGTGGAGGCTTGCGGACCAAGATCCGC	60
Db	1	TTGGCCACTCCCTCTCTGTGGCGCTGGCTCGCTCACTGAGCGCGGAGCAACAAAGGTGCC	60
OY	61	AGACGGCAGAGCTCTCTCTGTGGCGGCCACCGAGCGAGAGGCGCCAGAGAGGAGTG	120
Db	61	CGAGCGCCCGAGCTTTTGGCCCGCGCGGCTCTCAAGTGAACGAGCGCCGACGAGAGAGGAGTG	120
OY	121	GGCACTTCATCACTAGGGGT	141
Db	121	GCCAACTCATCACTTAGGGGT	141

61 AGACGGCAGAGCTCTGCTCTGCTGCGGCGCCCACCGAGCGAGCGGAGCGGCGGAGAGGGAGTGTG 120

QUESTIONS	ANSWERS
1. What is the purpose of the study?	To determine the effect of the new drug on the old drug.
2. What are the objectives of the study?	To determine the effect of the new drug on the old drug.
3. What are the hypotheses of the study?	To determine the effect of the new drug on the old drug.
4. What are the variables of the study?	To determine the effect of the new drug on the old drug.
5. What are the methods of the study?	To determine the effect of the new drug on the old drug.
6. What are the results of the study?	To determine the effect of the new drug on the old drug.
7. What are the conclusions of the study?	To determine the effect of the new drug on the old drug.
8. What are the recommendations of the study?	To determine the effect of the new drug on the old drug.
9. What are the limitations of the study?	To determine the effect of the new drug on the old drug.
10. What are the strengths of the study?	To determine the effect of the new drug on the old drug.

CURRENT APPLICATION NUMBER: US/10/240,198
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: 0009887.1
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 4675
TYPE: DNA
ORGANISM: adeno-associated virus 2
US-10-240-198-1

Query Match 75.1%; Score 107.4; DB 15; Length 4675;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTGCGGAGGCTTGCAGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTCGCTCACTGAGCGCGGACCAAGTCCGC 60
QY 61 AGAGGCGAGAGCTCTGCTCTGCGGCGCCCAACGAGGAGGCGGAGAGAGAGTG 120
DB 61 CGAGCGCCCGGCTTTGCGCGGCGCTCACTGAGGAGGAGGAGGAGGAGTG 120
QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GGCACTCCATCACTAGGGGT 141

RESULT 10
US-09-804-898-1
Sequence 1, Application US/09804898
Patent No. US20020045264A1
GENERAL INFORMATION:
APPLICANT: DURING, MATTHEW
APPLICANT: XIAO, WEIDONG
TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
FILE REFERENCE: 102182-14
CURRENT APPLICATION NUMBER: US/09/804,898
CURRENT FILING DATE: 2001-03-13
PRIOR APPLICATION NUMBER: 60/189,110
PRIOR FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin version 3.0
SEQ ID NO 1
LENGTH: 4679
TYPE: DNA
ORGANISM: adeno-associated virus 2
US-09-804-898-1

Query Match 75.1%; Score 107.4; DB 10; Length 4679;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTGCGGAGGCTTGCAGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTCACTGAGCGCGGACCAAGTCCGC 60
QY 61 AGAGGCGAGAGCTCTGCTCTGCGGCGCCCAACGAGGAGGCGGAGAGAGTG 120
DB 61 CGAGCGCCCGGCTTTGCGCGGCGCTCACTGAGGAGGAGGAGGAGGAGTG 120
QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GGCACTCCATCACTAGGGGT 141

RESULT 11
US-09-945-681-10
Sequence 10, Application US/09945681
Patent No. US20020064878A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITE DE NANTES

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
FILE REFERENCE: B4182A - UNIVERSITE DE NANTES
CURRENT APPLICATION NUMBER: US/09/945,681
CURRENT FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: PCT/EP 00/01854
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 10
LENGTH: 4679
TYPE: DNA
ORGANISM: adeno-associated virus 2
US-09-945-681-10

Query Match 75.1%; Score 107.4; DB 10; Length 4679;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTGCGGAGGCTTGCAGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTCACTGAGCGCGGACCAAGTCCGC 60
QY 61 AGAGGCGAGAGCTCTGCTCTGCGGCGCCCAACGAGGAGGCGGAGAGAGTG 120
DB 61 CGAGCGCCCGGCTTTGCGCGGCGCTCACTGAGGAGGAGGAGGAGGAGTG 120
QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GGCACTCCATCACTAGGGGT 141

RESULT 12
US-10-038-972A-12
Sequence 12, Application US/10038972A
Publication No. US20020192823A1
GENERAL INFORMATION:
APPLICANT: J. Bartlett
TITLE OF INVENTION: AAV VECTORS AND METHODS
FILE REFERENCE: 28335/3696US
CURRENT APPLICATION NUMBER: US/10/038,972A
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: US 60/260,124
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin version 3.1
SEQ ID NO 12
LENGTH: 4679
TYPE: DNA
ORGANISM: adeno-associated virus 2
US-10-038-972A-12

Query Match 75.1%; Score 107.4; DB 15; Length 4679;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTGCGGAGGCTTGCAGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGGCTCGCTCGCTCACTGAGCGCGGACCAAGTCCGC 60
QY 61 AGAGGCGAGAGCTCTGCTCTGCGGCGCCCAACGAGGAGGCGGAGAGAGTG 120
DB 61 CGAGCGCCCGGCTTTGCGCGGCGCTCACTGAGGAGGAGGAGGAGGAGTG 120
QY 121 GGCACTCCATCACTAGGGGT 141
DB 121 GGCACTCCATCACTAGGGGT 141

RESULT 13
US-10-077-294-1
Sequence 1, Application US/10077294
Publication No. US20020159979A1

GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/077,294
FILING DATE: 15-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/691,604
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. US20020159979A1and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-077-294-1
Query Match 75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGTCCGC 60
QY 61 AGACGGCAGAGCTGCTGCTGCGCGCGCCCAACGAGGAGGAGCGGCGAGAGGAGTG 120
DB 61 CGAGCGCGCGGCGCTTTGCGCGCGGCGCTCAGTGAAGGAGGAGCGGCGAGAGGAGTG 120
QY 121 GCGCACTCCATCACTAGGGGT 141
DB 121 GCCCACTCCATCACTAGGGGT 141
RESULT 14
US-10-163-886-1
Sequence 1, Application US/10163886
Publication No. US20020187129A1
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/163,886
FILING DATE: 04-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/292,703
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. US20020187129A1and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-163-886-1
Query Match 75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred. No. 1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TTGCCCACCTCCCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGGACCAAGTCCGC 60
DB 1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGTCCGC 60
QY 61 AGACGGCAGAGCTGCTGCTGCGCGCGCCCAACGAGGAGGAGCGGCGAGAGGAGTG 120
DB 61 CGAGCGCGCGGCGCTTTGCGCGCGGCGCTCAGTGAAGGAGGAGCGGCGAGAGGAGTG 120
QY 121 GCGCACTCCATCACTAGGGGT 141
DB 121 GCCCACTCCATCACTAGGGGT 141
RESULT 15
US-10-263-127-1
Sequence 1, Application US/10263127
Publication No. US20030082145A1
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/263,127
FILING DATE: 02-Oct-2002

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 212.5 Seconds

(without alignments)
10896.630 Million cell updates/sec

Title: US-09-807-802a-1_COPY_1_143
Perfect score: 143
Sequence: 1 ttgccacacccctctctgcg.....aactccatcactagggttaa 143

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba:*
2: em_estbum:*
3: em_estcin:*
4: em_estcun:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_estc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estcun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.2	26.7	763	12	BG490707 602520211
2	37.4	26.2	644	10	BB632576 BB632576
3	36.2	25.3	920	12	BG764368 BG764368
4	36.2	25.3	932	17	AL066742 Drosophila
5	35.8	25.0	604	13	BI531460 1024113H0
6	35.6	24.9	578	10	BB651469 BB651469

7	35.4	24.8	860	12	BF578052
8	35.4	24.8	1123	17	AG080476
9	35.2	24.6	2024	17	AG074763
10	35	24.5	844	17	CNS0052P
11	34.4	24.1	745	12	BG837739
12	34.4	24.1	890	17	AG127787
13	34	23.8	623	10	BB613613
14	34	23.8	639	10	BB651977
15	34	23.8	643	10	BB652152
16	34	23.8	657	10	BB651054
17	34	23.8	661	10	BB621081
18	34	23.8	795	17	AQ331299
19	33.8	23.6	932	17	CNS00720
20	33.8	23.6	1136	17	BM416106
21	33.8	23.6	1372	13	BI134462
22	33.8	23.5	984	12	BG809678
23	33.6	23.5	1020	17	CNS004YJ
24	33.6	23.5	1136	13	BM416106
25	33.6	23.5	265	14	TI4763
26	33.4	23.4	1009	17	CNS010EM
27	33.4	23.4	1284	13	BM547577
28	33.4	23.2	827	17	AZ198560
29	33.2	23.2	879	17	AG141413
30	33.2	23.2	925	17	CNS0091P
31	32.8	22.9	437	13	BM487026
32	32.6	22.8	727	13	BI454382
33	32.6	22.8	877	17	AG127610
34	32.6	22.8	1014	17	CNS0143J
35	32.6	22.8	337	10	BE110673
36	32.4	22.7	445	9	AL514067
37	32.4	22.7	749	12	BG837752
38	32.4	22.7	822	17	AG134283
39	32.4	22.7	1100	17	AG161988
40	32.4	22.5	529	12	BE967846
41	32.2	22.5	604	17	AQ575796
42	32.2	22.5	657	17	AG060145
43	32.2	22.5	664	17	AQ259032
44	32.2	22.5	692	17	AQ362958
45	32.2	22.5	692	17	AQ362958

ALIGNMENTS

RESULT 1
BG490707/c 763 bp mRNA linear EST 27-MAR-2001
LOCUS 60252021F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4638699 5'
DEFINITION mRNA sequence.
ACCESSION BG490707
VERSION BG490707.1 GI:13452217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 763)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCMT401 row: e column: 04
High quality sequence stop: 182.
Location/Qualifiers
1. 763

FEATURES
source

Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Cerninci, P., Shbata, Y., Hayatsu, N., Sugahara, Y., Shbata, K., Itoh
, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. *Genome Res.* 10 (10), 1611-1630 (2000)
Wegli, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BG764368	BG764368.1	GI:14075008	EST.	human.
				Homo sapiens
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	1 (bases 1 to 920)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. National Institutes of Health

1 (bases 1 to 920)
 NIH-MGC <http://imgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC).
 Unpublished (1999)
 Contact: Robert Strauberg, Ph. D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC/BCTD/DYK
 cDNA library Preparation: ling Hong/Rubin Laboratory
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
 plate: L1CM1720 row: d column: 16
 high quality sequence stop: 748.

FEATURES

Source

```

SOURCE
1. .920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4861407"
/clone_lib="NIH MGC 49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pT187; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
169 a 260 c 302 g 189 t

```

Query Match	25.3%	Score 36.2;	DB 12;	Length 920;
Best Local Similarity	56.2%	Pred. No. 19;		
Matches 68; Conservative	0;	Mismatches	53;	Indels 0; Gaps 0

[illegible]

LOCUS	CNS00720	932 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14B09 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

ACCESSION	AL066742	GI:4945205
VERSION	AL066742.1	
KEYWORDS	GSS.	
SOURCE	Drosophila melanogaster	
ORGANISM	Drosophila melanogaster	

REFERENCE	1 (bases 1 to 932)
AUTHORS	Genoscope.

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammosses in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

son

				/organism="Drosophila melanogaster"	
				/db_xref="taxon:7227"	
				/clone="BACR14B09"	
				/clone_1lb="RPC1-98"	
				/note="end : T7"	
BASE COUNT	155	a	202	c	241
				g	91
				t	243
				others	
ORIGIN					
Query Match	25.3%	Score 36.2,	DB 17,	Length 932,	
Best Local Similarity	36.3%	Pred. No. 19,			
Matches 41,	Conservative 32,	Mismatches 40,	Indels	0,	Gaps 0,
OY	4	CCGCACTCCTCTGCGCGCTCGCTCGCTGCGTGGGCGCTCGCGGCCAAGAGTCCGAGA	63		
Db	725	SSCGCSCTCCSSCCCCSSCGCCSSSGCGSSGCGSSSGCGSSCGCCSSCGGSGCGG	784		
OY	64	CGGCGAGACTCTGCTTGCCTGCGGCCCGACCGAGCGAGCGAGCGCGAGAGG	116		

RESULT 5

BI531460/C	BI531460	604 bp	mRNA	linear	EST 29-AUG-2001
LOCUS					
DEFINITION	1024113H09.y1 C. reinhardtii CC-1690, normalized, lambda zap II				
ACCESSION	BI531460	Chlamydomonas reinhardtii cDNA, mRNA sequence.			
VERSION	BI531460.1	GI:15372034			
KEYWORDS	EST.				
SOURCE	Chlamydomonas reinhardtii.				
ORGANISM	Chlamydomonas reinhardtii				
	Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;				
	Chlamydomonadaceae; Chlamydomonas.				
REFERENCE	1 (bases 1 to 604)				

Vascular Plants. Project: 1024b
JOURNAL
Unpublished (2001)
COMMENT
Contact: Charles Hauser
DCMB Box 91000

Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

```
/strain="CC-1690 wi
/db_xref="taxon:305
```

```

IIn
/note="Vector: pluescript II SK; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to

```


Qy 18 GCGGCTCGCTCGCTCGCTGGGGGCTGCGACCAAGTCCGACAGCGAGCTGCG 77
LOCUS Zm10_04h12.A Zm10_AAPC_ECORC_Fusarium_graminearum_corn_silk_Zea
Db 1047 GCGGNCAGCGCGGACCGAGCGGAGACCGCCCTCGNAGAGACGCGC 1106
Qy 78 TCTGCGGCGCCCGACCGAGCGAGCGCGAGAGAG 115
Db 1107 GCGCGCGCGCGGAGCGCGAGCGCGAGAGG 1144

RESULT 10
CNS0052P 844 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TERT end of BAC #
DEFINITION BAC11P16 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL056652
VERSION AL056652.1 GI:4932342
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 844)
Genoscope.
REFERENCE Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..844
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC11P16"
/clone_1lb="RPCT-98"
/note="End : TERT3"

BASE COUNT 261 a 112 c 92 g 35 t 344 others
ORIGIN

Query Match 24.5% Score 35; DB 17; Length 844;
Best Local Similarity 21.5% Pred. No. 38; Indels 0; Gaps 0;
Matches 26; Conservative 61; Mismatches 34;

Qy 3 GCCCACTCCCTCTGCGCGCTCGCTCGCTGGGGGCTGCGACCAAGTCCGAG 62
Db 490 SMCCSMCGSVSVACVSSGVSVAVGCGARVGSRRGAGSSSGRGSSSVSGV 549
Qy 63 ACAGGAGAGCTGTGCTGCGCGCGCGCGCGCGAGCGAGCGCGAGAGAGTGG 122
Db 550 SSSSSVWGACASSSVSCBSVSVSSGBVSRGCGRCVGGVGGSSSVSSGSS 609
Qy 123 C 123
Db 610 S 610

RESULT 11

BG837739/c 745 bp mRNA linear EST 25-MAY-2001
LOCUS Zm10_04h12.A Zm10_AAPC_ECORC_Fusarium_graminearum_corn_silk_Zea
DEFINITION mays_cDNA clone Zm10_04h12, mRNA sequence.
ACCESSION BG837739
VERSION BG837739.1 GI:14204062
KEYWORDS EST.
SOURCE Zea mays:
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 745)
REFERENCE Harris,L.J., Balcerzak,M., Allard,S., Saperio,A., Couroux,P., De
Moore,A., Hattori,J.I., Queller,T., Robert,L.S., Singh,J.A., Sprott
J.D. and Tinker,N.A. Expressed Sequence Tags from Maize Silk Six Hours After Silk
Channel Inoculation with Fusarium graminearum
Unpublished (2001)
CONTACT: Harris, Linda J.
Eastern Cereal and Agri-food Research Centre
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris@em.agr.ca
Location/Qualifiers
1..745
/organism="Zea mays"
/cultivar="CO388"
/db_xref="taxon:4577"
/clone="Zm10_04h12"
/clone_1lb="Zm10_AAPC_ECORC_Fusarium_graminearum_corn_silk
" /tissue_type="Silk"
/dev_stage="4-5 days post-silk emergence"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Field-grown corn was silk channel-inoculated
in the morning (-10 am) with 1 ml of a macroconidial
suspension (500,000 spores/ml) of Fusarium graminearum and
silk channels were collected and immediately frozen in
liquid nitrogen 6 hours later. RNA was extracted from
silk tissue between 1 cm below and above the inoculation
point in the silk channel. RNA from five silk channels was
pooled."

BASE COUNT 134 a 211 c 263 g 137 t
ORIGIN

Query Match 24.1% Score 34.4; DB 12; Length 745;
Best Local Similarity 63.1% Pred. No. 52;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 56 TCCGACAGCGAGAGCTGTGCTGCGCGCGCGCGCGAGCGAGCGCGAGAGG 115
Db 727 TCTACTACCTGTGACCTCGCATTCCTCAACAGCCGAGCGAGCTAGAGAGCGG 668
Qy 116 GAGTGGCACTCCATCATAGG 139
Db 667 CGGAGGACCTCTCTCTCAAG 644

RESULT 12
AG127787 890 bp DNA linear GSS 04-NOV-2001
LOCUS Pan trogodytes DNA, clone: PTB-138W02.F, genomic survey sequence.
DEFINITION AG127787
ACCESSION AG127787
VERSION AG127787.1 GI:1665952
KEYWORDS GSS.
SOURCE Pan trogodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
BAC library clone:PTB-138W02.F.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Kondo, S., Shinagawa, A., Sato, T., Kiyosawa, H., Yamazaki, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawaj, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1. 643
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330049B17"
/clone_lib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genetic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGGAGAGAGAGATTCGACCAAGAGCGCTTTTTTTTTTTTAA 3']₁. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA was then through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGGAGAGAGATTCGACCAAGATTAAATTAATATATATCCCCCCCC 3']₁.

BASE COUNT	102 a	242 c	161 g	138 t
ORIGIN				

Query Match	23.8%	Score 34	DB 10	Length 643
Best Local Similarity	4.4%	Pred. No. 63		
Matches	52	Conservative	0	Mismatches 30; Indels 0; Gaps 0;
QY	11	CCTCTCTGCGGCGCTCTGCGCTGCGGCGCTGCGGCGCAAGATCGCGAGAGCGGAGA	70	
DB	58	CCTCTCGGCGCCTCGGCGAGTCTGCGCGCGCGCGGCGAGAGACACGACGACGACCATG	117	
QY	71	GCTCTGCTCTGCGGCGCCGACCC	92	
DB	118	CCTCTACTTTGACCGTCCACCC	139	

Search completed: July 20, 2003, 15:27:49
Job time : 216.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 ; Search time 51.1932 Seconds

(without alignments)
1915.734 Million cell updates/sec

Title: US-09-807-802a-13

Sequence: 1 MAADGYLPDMLIEDNLSEGR.....NGLYTEPRPIGTRVLRPL 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3989	100.0	736	AAV71167	Adeno-associated v
2	3963	99.3	736	AA859847	AAV6 capsid protei
3	3511	88.0	736	AA859846	AAV3 capsid protei
4	3494	87.6	736	AA859845	AAV4 capsid protei
5	3402.5	85.3	735	AA65792	Adeno-associated v
6	3402.5	85.3	735	AA851508	Adeno-associated v
7	3402.5	85.3	735	AA859844	AAV2 capsid protei
8	3376	84.6	734	AA850326	Adeno-associated v
9	3251	81.5	599	AAV71168	Adeno-associated v
10	2906	72.9	534	AAV71169	Adeno-associated v

11	2759.5	69.2	598	22	AA65793	Adeno-associated v
12	2759.5	69.2	598	22	AA851509	Adeno-associated v
13	2486.5	62.3	734	19	AAW46308	AAV4 VPI capsid pr
14	2481.5	62.2	533	22	AA65794	Adeno-associated v
15	2477.5	62.1	533	22	AA851510	Adeno-associated v
16	2220	55.7	724	21	AAV58160	Adeno-associated v
17	2220	55.7	724	23	AAU11405	Adeno-associated v
18	2195.5	55.0	736	19	AAU19000	Duck parvovirus ca
19	1830.5	54.4	732	16	AA85385	Barbary duck parvo
20	1830.5	45.9	598	19	AAW46313	AAV4 VP2 coat prot
21	1700.5	42.6	588	21	AAV58161	Adeno associated v
22	1700.5	42.6	588	23	AAU11406	Adeno-associated v
23	1690.5	42.4	544	19	AAW46314	AAV4 VP3 coat prot
24	1676	42.0	534	16	AA85386	Barbary duck parvo
25	1665	41.7	532	23	AAV58162	Adeno associated v
26	1665	41.7	532	23	AAU11407	Adeno-associated v
27	1410	35.3	756	21	AAV71231	Capsid protein enc
28	921	23.1	202	23	AAU83481	Adeno-associated v
29	606.5	15.2	781	20	AAV23227	Erythrovirus V9 VP
30	605.5	15.2	781	16	AAW08986	Human parvovirus V
31	490	12.3	554	16	AAW08987	Human parvovirus V
32	490	12.3	554	20	AAV23230	Erythrovirus V9 VP
33	476.5	11.9	543	12	AAU13405	Parvo virus B19 VP
34	425.5	10.7	472	5	AAU0675	Sequence encoded b
35	417.5	10.5	620	22	AAW50111	Feline parvovirus
36	318	8.0	59	23	AAU83456	Parvovirus VP1up c
37	310	7.8	264	12	AAU13407	Parvo virus B19 PA
38	303	7.6	59	23	AAU83454	Parvovirus VP1up c
39	301	7.5	59	23	AAU83453	Parvovirus VP1up c
40	297	7.4	59	23	AAU83452	Parvovirus VP1up c
41	295.5	7.4	370	12	AAU13406	Parvo virus B19 PA
42	282	7.1	686	5	AAU40068	Sequence of a porc
43	262.5	6.6	172	23	AAU83476	Porcine parvovirus
44	255	6.4	59	23	AAU83448	Parvovirus VP1up c
45	252	6.3	59	23	AAU83455	Parvovirus VP1up c

ALIGNMENTS

RESULT 1	
AAV71167	
AAV71167 standard; Protein; 736 AA.	
XX	
AC	AAV71167;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Adeno-associated virus serotype 1 capsid protein VPI.
XX	
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KM	cap protein; recombinant viral vector; gene delivery; gene therapy;
XX	
KW	vaccine; transgene; VPI.
XX	
OS	Adeno associated virus serotype 1.
XX	
PN	MO200028061-AA.
XX	
PD	18-MAY-2000.
XX	
PF	02-NOV-1999; 99MO-US25694.
XX	
PR	05-NOV-1998; 98US-0107114.
XX	
PA	(VPE-) UNITV PENNSYLVANIA.
XX	
PI	Wilson JM, Xiao W;
XX	
DR	WPI: 2000-376571/32.
XX	
DR	N-PSDB; AAD00772, AAD00777.
XX	
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for preparation of medicament for delivery of a transgene to a host


```

Db 241 TTSTRWALPTYNHLYKOISSASTAGSNHFGSTPMGYDFRPHCHSPRMOQL 300
Qy 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTIANLSTVQVPSSEYOLPYVLGSAHQ 360
Db 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTIANLSTVQVPSSEYOLPYVLGSAHQ 360
Qy 361 GCLPPPADVFMTPOGYTLTNNGSAQVGRSSFCLEYPSPQMLRTGNFTSYTEEDVP 420
Db 361 GCLPPPADVFMTPOGYTLTNNGSAQVGRSSFCLEYPSPQMLRTGNFTSYTEEDVP 420
Qy 421 FHSSVYASOSLDRLNPLIDQVLYLNRTQNGSAGNKLDFRSRGSFAGMSYQPKXWLP 480
Db 421 FHSSVYASOSLDRLNPLIDQVLYLNRTQNGSAGNKLDFRSRGSFAGMSYQPKXWLP 480
Qy 481 GPCYRQORVSKTIDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDXKDFPMSGV 540
Db 481 GPCYRQORVSKTIDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDXKDFPMSGV 540
Qy 541 MFGKESAGASNTALDNVMTDEEIKATNPVATERGTAVANFQSSSTDPATGDYHAMG 600
Db 541 MFGKESAGASNTALDNVMTDEEIKATNPVATERGTAVANFQSSSTDPATGDYHAMG 600
Qy 601 ALPGMWODRDVYLQGPIMAKIPHTDGHFSPPLMGFGJLKNPPQILLIKNTVPANPPA 660
Db 601 ALPGMWODRDVYLQGPIMAKIPHTDGHFSPPLMGFGJLKNPPQILLIKNTVPANPPA 660
Qy 661 EFSATKFASTITQYSTQGVSEIEMELQKENSKRANPEVOYTSNAYKSNVDTVDNNGL 720
Db 661 EFSATKFASTITQYSTQGVSEIEMELQKENSKRANPEVOYTSNAYKSNVDTVDNNGL 720
Qy 721 YTEPRPIGRYLRPL 736
Db 721 YTEPRPIGRYLRPL 736

RESULT 3
AAB59846
ID AAB59846 standard; Protein; 736 AA.
XX
AC AAB59846;
XX
XX 28-MAR-2001 (first entry)
XX
DE AAV3B capsid protein VP1.
XX
XX AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX atherosclerosis; sickle cell anaemia; thalassemia;
XX blood clotting disorder; diabetes; capsid protein VP1.
XX
OS Adeno associated virus.
XX
XX US6156303-A.
XX
XX 05-DEC-2000.
XX
XX 11-JUN-1997; 97US-0873168.
XX
XX 11-JUN-1997; 97US-0873168.
XX (UNIW ) UNIV WASHINGTON.
XX
XX Russell DW, Rutledge EA;
XX
XX WPI; 2001-060164/07.
XX
XX Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX
XX Disclosure; Fig 2; 50pp; English.
XX
XX The present invention relates to adeno-associated virus serotypes. The

```

```

CC present sequence is capsid protein VP1 of one such serotype (AAV3B).
CC AAV3B can be used to construct AAV viral vectors for use in gene therapy
CC for a range of disorders; cystic fibrosis, cancer, AIDS, atherosclerosis,
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
CC The AAV viral vectors have increased transduction efficiency of a
CC particular host cell as the AAV virion containing the AAV vector genome
CC can be modified to express a capsid protein of an AAV serotype that
CC transduces the selected host cell.
XX
XX Sequence 736 AA;
XX
Qy 1 MAADGYPMLBNDLSEGRWMDLKPAPKPKANQKDDGRLVLPQYKYLGPNGID 60
Db 1 MAADGYPMLBNDLSEGRWMDLKPAPKPKANQKDDGRLVLPQYKYLGPNGID 60
Qy 61 KGPVNAADAALHDKAVDQOLKAGDNPLYLNHADAFOERLQEDTSFGGNLGRAVHQ 120
Db 61 KGPVNAADAALHDKAVDQOLKAGDNPLYLNHADAFOERLQEDTSFGGNLGRAVHQ 120
Qy 121 AKKRVLPEGLIVEBAGAKTAPKKRPVQSPQEPDSSSGIGKTGOQPAKRLNFGQTGSE 180
Db 121 AKKRVLPEGLIVEBAGAKTAPKKRPVQSPQEPDSSSGIGKTGOQPAKRLNFGQTGSE 180
Qy 181 SVDPDPLCEPPATPAVGPPTMASGGAPMADNNEGADGVGNASGNHCDSTLGDRIY 240
Db 181 SVDPDPLCEPPATPAVGPPTMASGGAPMADNNEGADGVGNASGNHCDSTLGDRIY 240
Qy 241 TTSTRWALPTYNHLYKOISSASTAGSNHFGSTPMGYDFRPHCHSPRMOQL 300
Db 241 TTSTRWALPTYNHLYKOISSASTAGSNHFGSTPMGYDFRPHCHSPRMOQL 300
Qy 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTIANLSTVQVPSSEYOLPYVLGSAHQ 360
Db 301 INNWGFRPKRLNFKLFNIQVKEVTNDGVTIANLSTVQVPSSEYOLPYVLGSAHQ 360
Qy 361 GCLPPPADVFMTPOGYTLTNNGSAQVGRSSFCLEYPSPQMLRTGNFTSYTEEDVP 420
Db 361 GCLPPPADVFMTPOGYTLTNNGSAQVGRSSFCLEYPSPQMLRTGNFTSYTEEDVP 420
Qy 421 FHSSVYASOSLDRLNPLIDQVLYLNRTQNGSAGNKLDFRSRGSFAGMSYQPKXWLP 480
Db 421 FHSSVYASOSLDRLNPLIDQVLYLNRTQNGSAGNKLDFRSRGSFAGMSYQPKXWLP 480
Qy 480 PGPYRQORVSKTIDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDXKDFPMSGV 540
Db 480 PGPYRQORVSKTIDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDXKDFPMSGV 540
Qy 540 VMI FGKESAGASNTALDNVMTDEEIKATNPVATERGTAVANFQSSSTDPATGDYHAM 599
Db 540 VMI FGKESAGASNTALDNVMTDEEIKATNPVATERGTAVANFQSSSTDPATGDYHAM 599
Qy 600 GALPGMWODRDVYLQGPIMAKIPHTDGHFSPPLMGFGJLKNPPQILLIKNTVPANPP 659
Db 600 GALPGMWODRDVYLQGPIMAKIPHTDGHFSPPLMGFGJLKNPPQILLIKNTVPANPP 659
Qy 660 AEFSAKFASTITQYSTQGVSEIEMELQKENSKRANPEVOYTSNAYKSNVDTVDNNG 719
Db 660 AEFSAKFASTITQYSTQGVSEIEMELQKENSKRANPEVOYTSNAYKSNVDTVDNNG 719
Qy 720 LYTEPRPIGRYLRPL 736
Db 720 LYTEPRPIGRYLRPL 736

RESULT 4
AAB59845
ID AAB59845 standard; Protein; 736 AA.
XX
XX AAB59845;

```

XX 28-MAR-2001 (first entry)
 XX AAV3A capsid protein VP1.
 XX AAV3A; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 XX atherosclerosis; sickle cell anaemia; thalassemia;
 XX blood clotting disorder; diabetes; capsid protein VP1.
 XX Adeno associated virus.
 XX US6156303-A.
 XX 05-DEC-2000.
 PD 11-JUN-1997; 97US-0873168.
 XX 11-JUN-1997; 97US-0873168.
 PR 11-JUN-1997; 97US-0873168.
 XX (UNIV) UNIV WASHINGTON.
 PA Russell DW, Rutledge EA;
 XX WPI; 2001-060164/07.
 XX Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalassemia and diabetes
 XX Disclosure; Fig 2; 50pp; English.
 XX The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is capsid protein VP1 of one such serotype (AAV3A).
 CC AAV3A can be used to construct AAV viral vectors for use in gene therapy
 CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.
 CC SQ Sequence 736 AA;
 Query Match 87.6%; Score 3494; DB 22; Length 736;
 Best Local Similarity 86.4%; Pred. No. 5,66-272;
 Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;
 QY 1 MAADGYLPMLEEDNLSEGIKREWDLPKQAPKPKANQKQDGRGLVLPGYKYTLGPFNGLD 60
 DB 1 MAADGYLPMLEEDNLSEGIKREWDLPKQAPKPKANQKQDGRGLVLPGYKYTLGPFNGLD 60
 QY 61 KGEFVNADAAALAHDKAYDQOLKAGDNPKYKYNHAAAEFQRIQEBTSPFGNIGRAVFO 120
 DB 61 KGEFVNADAAALAHDKAYDQOLKAGDNPKYKYNHAAAEFQRIQEBTSPFGNIGRAVFO 120
 QY 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPSSSGIGTGOQPAKKRLNFGQTDSE 180
 DB 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPSSSGIGTGOQPAKKRLNFGQTDSE 180
 QY 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPSSSGIGTGOQPAKKRLNFGQTDSE 180
 DB 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPSSSGIGTGOQPAKKRLNFGQTDSE 180
 QY 181 SVDPQPLGEBEPATPAVGPPTMASGGAPMADNNEGADGVGNASGMWCHDSTWLGDRVI 240
 DB 181 SVDPQPLGEBEPATPAVGPPTMASGGAPMADNNEGADGVGNASGMWCHDSTWLGDRVI 240
 QY 181 SVDPQPLGEBEPATPAVGPPTMASGGAPMADNNEGADGVGNASGMWCHDSTWLGDRVI 240
 DB 181 SVDPQPLGEBEPATPAVGPPTMASGGAPMADNNEGADGVGNASGMWCHDSTWLGDRVI 240
 QY 241 TTSTRTALPTNNHLYKQISSASTGASNDNHFGYSPMKVFPFNNFCHFSRDMQRL 300
 DB 241 TTSTRTALPTNNHLYKQISSASTGASNDNHFGYSPMKVFPFNNFCHFSRDMQRL 300
 QY 241 TTSTRTALPTNNHLYKQISSASTGASNDNHFGYSPMKVFPFNNFCHFSRDMQRL 300
 DB 241 TTSTRTALPTNNHLYKQISSASTGASNDNHFGYSPMKVFPFNNFCHFSRDMQRL 300
 QY 301 INNWGFRPKLNFNIQVEKVTNDGVTTIANNLSTVQVPSDEYQLPYLGSAHQ 360
 DB 301 INNWGFRPKLNFNIQVEKVTNDGVTTIANNLSTVQVPSDEYQLPYLGSAHQ 360
 QY 300 INNWGFRPKLNFNIQVEKVTNDGVTTIANNLSTVQVPSDEYQLPYLGSAHQ 359
 DB 300 INNWGFRPKLNFNIQVEKVTNDGVTTIANNLSTVQVPSDEYQLPYLGSAHQ 359
 QY 361 GCLPPPADVFMIDQGYLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTSYTFEEVP 420
 DB 361 GCLPPPADVFMIDQGYLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTSYTFEEVP 420

DB 360 GCLPPPADVFMIDQGYLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTSYTFEEVP 419
 QY 421 FHSSVAHSGSLDPLNPLIDQYLYNLRPC-NOSGSAQNKDILFSGSPAGMSVQPKML 479
 DB 420 FHSSVAHSGSLDPLNPLIDQYLYNLRPC-NOSGSAQNKDILFSGSPAGMSVQPKML 479
 QY 480 PGCCYQOQRSKTKTNNNSNFTWTGASKYKLNIGRESIINPGTAMASHKDEDEKFFPMG 539
 DB 480 PGCCYQOQRSKTKTNNNSNFTWTGASKYKLNIGRESIINPGTAMASHKDEDEKFFPMG 539
 QY 540 VMI FGKESAGASTALDNWITDEBEIKATNPVATRFPTVAVNFSSSDPATGVHAM 599
 DB 540 VMI FGKESAGASTALDNWITDEBEIKATNPVATRFPTVAVNFSSSDPATGVHAM 599
 QY 540 NLI FGKESAGASTALDNWITDEBEIKATNPVATRFPTVAVNFSSSDPATGVHAM 599
 DB 540 NLI FGKESAGASTALDNWITDEBEIKATNPVATRFPTVAVNFSSSDPATGVHAM 599
 QY 600 GALPGWMDRDVYLGPIWAKIPIHTDGHFHPSPILMGREGKPKPPQIILKNTVPANRP 659
 DB 600 GALPGWMDRDVYLGPIWAKIPIHTDGHFHPSPILMGREGKPKPPQIILKNTVPANRP 659
 QY 660 AEFSAATKFAFTIYQSTGVSVIEWELQENSKRNPEVQYTSNYAKSANVDFTVNNNG 719
 DB 660 AEFSAATKFAFTIYQSTGVSVIEWELQENSKRNPEVQYTSNYAKSANVDFTVNNNG 719
 QY 660 TTFSPAKFASFTIYQSTGVSVIEWELQENSKRNPEVQYTSNYAKSANVDFTVNNNG 719
 DB 660 TTFSPAKFASFTIYQSTGVSVIEWELQENSKRNPEVQYTSNYAKSANVDFTVNNNG 719
 QY 720 LYTEPRPIGTRYLTRPL 736
 DB 720 LYTEPRPIGTRYLTRPL 736
 QY 720 VYSEPRPIGTRYLTRPL 736
 DB 720 VYSEPRPIGTRYLTRPL 736
 RESULT 5
 AAG65792
 ID AAG65792 standard; Protein; 735 AA.
 XX AAG65792;
 AC AAG65792;
 DT 11-FEB-2002 (first entry)
 XX Adeno-associated virus 2 (AAV-2) major coat protein VP1.
 DE Adeno-associated virus 2 (AAV-2) major coat protein VP1.
 XX Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
 KM inverted terminal repeat; neurotropic; neuroprotective; antineoplastic;
 KM antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
 KM major coat protein; AAV-2; VP1.
 XX Adeno-associated virus 2.
 OS WO200168888-A2.
 XX WO200168888-A2.
 PD 20-SEP-2001.
 XX 13-MAR-2001; 2001WO-US07927.
 PF 13-MAR-2001; 2001WO-US07927.
 XX 14-MAR-2000; 2000US-189110P.
 PR 14-MAR-2000; 2000US-189110P.
 XX (NEUR-) NEUROLOGIX INC.
 PA Xiao W, During MJ;
 XX WPI; 2001-596912/67.
 DB N-PDB; AAI66974.
 DR N-PDB; AAI66974.
 XX The invention provides a recombinant viral vector (RVV) comprising a
 CC chimeric capsid (I) having at least one non-native amino acid sequence,
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),
 CC or their combination, and a transgene flanked 5' and 3' by inverted
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their
 CC combination. The RVV is useful for improving gene therapy in a subject,
 CC with a disorder, and for increasing the efficiency of entry into a cell,

CC which involves producing (I) encapsulating a viral vector, and contacting
 CC a cell with RVV having (I) such that (I) binds to an attachment site on
 CC the cell surface and permits the vector to enter the cell efficiently.
 CC A pharmaceutical composition comprising RVV with (I) containing a
 CC transgene sequence associated with a disease or a disorder such that
 CC expression of the transgene would result in amelioration of the disease
 CC or disorder such as inherited neurological and metabolic diseases e.g.,
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,
 CC retinoblastoma and various types of neoplastic cells which include
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,
 CC sarcomas, leukemias and lymphoma. The present sequence represents the
 CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences
 CC are used in the construction of a chimeric vector.

CC Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;

Best Local Similarity 83.3%; Pred. No. 1.3e-264; Mismatches 71; Indels 1; Gaps 1;

Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60
 1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60
 DB 1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60
 QY 61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120
 61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120
 DB 61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120
 QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEGSPQEPDSSSGIKTGQOPAKKRLNFGQTGDS 180
 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEGSPQEPDSSSGIKTGQOPAKKRLNFGQTGDS 180
 DB 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEGSPQEPDSSSGIKTGQOPAKKRLNFGQTGDS 180
 QY 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNESADGVAGNAGNHCDSITWLDGRVI 240
 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNESADGVAGNAGNHCDSITWLDGRVI 240
 DB 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNESADGVAGNAGNHCDSITWLDGRVI 240
 QY 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 DB 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 QY 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 DB 241 TTSRTWALPTYNHNLKYOISSASTGASNDNHFGYSTPMGYDPFNRFHCHFSPRDQRL 300
 QY 301 INNMGFRPKRLNFKLNIQVEKVTNDGVTIANNLSTVQVPSDEYQLPYVLSAHQ 360
 301 INNMGFRPKRLNFKLNIQVEKVTNDGVTIANNLSTVQVPSDEYQLPYVLSAHQ 360
 DB 301 INNMGFRPKRLNFKLNIQVEKVTNDGVTIANNLSTVQVPSDEYQLPYVLSAHQ 360
 QY 361 GCLPFPFADVFMIPOYGYLTLLNNGSQAVGRSSFFCYCLEFSPSOMLRTGNNTFFSYTFEEVP 420
 361 GCLPFPFADVFMIPOYGYLTLLNNGSQAVGRSSFFCYCLEFSPSOMLRTGNNTFFSYTFEEVP 420
 DB 361 GCLPFPFADVFMIPOYGYLTLLNNGSQAVGRSSFFCYCLEFSPSOMLRTGNNTFFSYTFEEVP 420
 QY 421 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 480
 421 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 480
 DB 421 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 480
 QY 420 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 479
 420 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 479
 DB 420 FHSVYASQGLDRLMNLIDQYLYLNRTONQSSAONKDLPSRGSAPGMSVQPKWLP 479
 QY 481 GPCYRQORVSKTDDNNNSFTWGSKNLNGEESTINPGTMASSKDEDEKFFPFGSV 540
 481 GPCYRQORVSKTDDNNNSFTWGSKNLNGEESTINPGTMASSKDEDEKFFPFGSV 540
 DB 481 GPCYRQORVSKTDDNNNSFTWGSKNLNGEESTINPGTMASSKDEDEKFFPFGSV 540
 QY 541 MIRGESAGASNTALDVMITDEBEIKATNPVATERGTVAVNFQSSSTPATGDVYAMG 600
 541 MIRGESAGASNTALDVMITDEBEIKATNPVATERGTVAVNFQSSSTPATGDVYAMG 600
 DB 541 MIRGESAGASNTALDVMITDEBEIKATNPVATERGTVAVNFQSSSTPATGDVYAMG 600
 QY 540 LIRGKQSSSEKTNVDIEKMTTDEBEIRTTNPVATEQGSVSTNLQDRNQALADAVTQGS 599
 540 LIRGKQSSSEKTNVDIEKMTTDEBEIRTTNPVATEQGSVSTNLQDRNQALADAVTQGS 599
 DB 540 LIRGKQSSSEKTNVDIEKMTTDEBEIRTTNPVATEQGSVSTNLQDRNQALADAVTQGS 599
 QY 601 ALPGMWQODRDVYLQGIWAKIPIHTDGHFSPSPIMSGFGLKNPPQILLINTPVPANPA 660
 601 ALPGMWQODRDVYLQGIWAKIPIHTDGHFSPSPIMSGFGLKNPPQILLINTPVPANPA 660
 DB 601 ALPGMWQODRDVYLQGIWAKIPIHTDGHFSPSPIMSGFGLKNPPQILLINTPVPANPA 660
 QY 661 EBSATKASRTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGL 720
 661 EBSATKASRTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGL 720
 DB 661 EBSATKASRTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGL 720
 QY 660 TFEAAKASFTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGV 719
 660 TFEAAKASFTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGV 719
 DB 660 TFEAAKASFTIYOSTGQVVEIEMELQKENSRRWNEVOYTSYASAVDFTVDMNGV 719
 QY 721 YTEPRPIGTGYLTRPL 736
 721 YTEPRPIGTGYLTRPL 736
 DB 721 YTEPRPIGTGYLTRPL 736

DB 720 YTEPRPIGTGYLTRPL 735

RESULT 6

AAAM51508 standard; protein; 735 AA.

AAAM51508;

02-JAN-2002 (first entry)

Adeno-associated virus VP1 capsid protein.

Adeno-associated virus; AAV; VP1; capsid; virus-like particle;

nuclear localisation signal; VP1.

Adeno associated virus.

JP2001169777-A.

26-JUN-2001.

30-JUL-1999; 99JP-0249140.

30-JUL-1999; 99JP-0249140.

(HAND/) HANDA H.

WPI; 2001-599854/68.

New virus-like particles from VP3 capsid protein of adeno-associated

PT virus, comprise a peptide containing a nucleus-shifting signal

PS connected to its N-terminal -

Disclosure; Page 10-13; 33pp; Japanese.

The present sequence is provided in a specification relating to a

CC virus-like particle-forming protein, and to a peptide containing a

CC nuclear-shifting signal at its N-terminus that can form virus-like

CC particles by shifting to the nucleus of the animal cell in which it

CC is expressed. The method is used for forming virus-like particles

CC from the VP3 capsid protein of adeno-associated virus (AAV).

CC Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;

Best Local Similarity 83.3%; Pred. No. 1.3e-264; Mismatches 71; Indels 1; Gaps 1;

Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60

1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60

DB 1 MAADGYLPWLEBNTLSEGIEMWDLKPGAPKPKANQOKODDGRGLVPGYKYLGPFGNGLD 60

QY 61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120

61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120

DB 61 KGEPVNAADAAALHDKAYDOOLKAGDNPLYRNHADAERFOERLOEDTSFGNIGRAVFO 120

QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEGSPQEPDSSSGIKTGQOPAKKRLNFGQTGDS 180

121 AKKRVLEPLGLVEEGAKTAPGKKRPVEGSPQEPDSSSGIKTGQOPAKKRLNFGQTGDS 180

QY 361 GCLPPFADVFMIDYGYLTLLNNGSOAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVP 420
 DB 360 GCLPPFADVFMVQYGYLTLLNNGSOAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVP 419
 QY 421 FHSSVAHSQSLDRLMNPILIDQYLYLNTONOGSQAONKDLIFSRGSPAGMSVOPKMWLP 480
 DB 420 FHSSVAHSQSLDRLMNPILIDQYLYLNTONOGSQAONKDLIFSRGSPAGMSVOPKMWLP 479
 QY 481 GPCYRQORVSKTKDNNNSFTWGSQVYIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 540
 DB 480 GPCYRQORVSKTKDNNNSFTWGSQVYIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 539
 QY 541 MIFKESAGASNTALDNNMTDEBEIKATNPVATERFQTVAVNFQSSSTDATGADVAMG 600
 DB 540 LIFGKQSEKTNVDIEKMTIDEERIITNPVATEQGSVSTNLQGRNRQAATADVMTQG 599
 QY 601 ALPGMWODRDVYLQGPIMAKIPHTDGHFHPSPLMGFGGLKMPPOILLIKNTFVPANPRA 660
 DB 600 VLPQMWODRDVYLQGPIMAKIPHTDGHFHPSPLMGFGGLKMPPOILLIKNTFVPANPRA 659
 QY 661 EFSATKFPASFTIOTSTGQVSVIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 720
 DB 660 TFSAAKFPASFTIOTSTGQVSVIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 719
 QY 721 YTEPRPIGTIRYLTRPL 736
 DB 720 YSEPRPIGTIRYLTRNL 735

RESULT 7
 AAB59844
 ID AAB59844 standard; Protein; 735 AA.

XX AC AAB59844;
 XX DT 28-MAR-2001 (first entry)
 XX DE AAV2 capsid protein VPI.
 XX AA AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 KW atherosclerosis; sickle cell anaemia; thalassemia;
 KW blood clotting disorder; diabetes; capsid protein VPI.
 XX OS Adeno associated virus.
 XX PN US6156303-A.
 XX PD 05-DEC-2000.
 XX PF 11-JUN-1997; 97US-0873168.
 XX PR 11-JUN-1997; 97US-0873168.
 XX PA (UNIT) UNIV WASHINGTON.
 XX PI Ruesell DW, Rutledge EA;
 XX DR WPI; 2001-060164/07.
 XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
 XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 XX syndrome, sickle cell anemia, thalassemia and diabetes
 XX PS Claim 7; Fig 2; 50pp; English.
 CC The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is capsid protein VPI of one such serotype (AAV2). AAV2
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that

CC transduces the selected host cell.

XX Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;
 Best Local Similarity 83.3%; Pred. No. 1.3e-264;
 Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDMLLENTLSBGTFFEMMDLKPAGKPKRANQKODGGGLVPGYKYLGPFGGLD 60
 DB 1 MAADGYLPDMLLENTLSBGTFFEMMDLKPAGKPKRANQKODGGGLVPGYKYLGPFGGLD 60
 QY 61 KGBPVNAADAAALAHDKAVYQQLKAGDNPYLRYNHADAERQERLOEDTSPGMLAGAVFQ 120
 DB 61 KGBPVNAADAAALAHDKAVYQQLKAGDNPYLRYNHADAERQERLOEDTSPGMLAGAVFQ 120
 QY 121 AKRVLVEPLGLVEGAKTAPGKKRPVYQSEQOEEDSSGIGTKTQOPAKKRLNFGOTGDE 180
 DB 121 AKRVLVEPLGLVEGAKTAPGKKRPVYQSEQOEEDSSGIGTKTQOPAKKRLNFGOTGDE 180
 QY 181 SVDPDPOPLGEPPTPAVGEFTTMAAGGAPMADNNGADGVNAGSNMHCDSITWLGDRVY 240
 DB 181 SVDPDPOPLGEPPTPAVGEFTTMAAGGAPMADNNGADGVNAGSNMHCDSITWLGDRVY 240
 QY 241 TTSRTWALPTYNHLYKOISSASTGASNDNHYFGYSTPWGYPDFNRFHCHFSPRDMORL 300
 DB 241 TTSRTWALPTYNHLYKOISSASTGASNDNHYFGYSTPWGYPDFNRFHCHFSPRDMORL 299
 QY 301 INNMGFRPKRLNFKLPNIOVKEVTNDGVTIANNLSTVQVFSQSEYOLPYVLSAQ 360
 DB 300 INNMGFRPKRLNFKLPNIOVKEVTNDGVTIANNLSTVQVFSQSEYOLPYVLSAQ 359
 QY 361 GCLPPFADVFMIDYGYLTLLNNGSOAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVP 420
 DB 360 GCLPPFADVFMIDYGYLTLLNNGSOAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVP 419
 QY 421 FHSSVAHSQSLDRLMNPILIDQYLYLNTONOGSQAONKDLIFSRGSPAGMSVOPKMWLP 480
 DB 420 FHSSVAHSQSLDRLMNPILIDQYLYLNTONOGSQAONKDLIFSRGSPAGMSVOPKMWLP 479
 QY 481 GPCYRQORVSKTKDNNNSFTWGSQVYIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 540
 DB 480 GPCYRQORVSKTKDNNNSFTWGSQVYIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 539
 QY 541 MIFKESAGASNTALDNNMTDEBEIKATNPVATERFQTVAVNFQSSSTDATGADVAMG 600
 DB 540 LIFGKQSEKTNVDIEKMTIDEERIITNPVATEQGSVSTNLQGRNRQAATADVMTQG 599
 QY 601 ALPGMWODRDVYLQGPIMAKIPHTDGHFHPSPLMGFGGLKMPPOILLIKNTFVPANPRA 660
 DB 600 VLPQMWODRDVYLQGPIMAKIPHTDGHFHPSPLMGFGGLKMPPOILLIKNTFVPANPRA 659
 QY 661 EFSATKFPASFTIOTSTGQVSVIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 720
 DB 660 TFSAAKFPASFTIOTSTGQVSVIEIEMELQKENSKRMPPEVOYTSNYSKANDVFTVDDNG 719
 QY 721 YTEPRPIGTIRYLTRPL 736
 DB 720 YSEPRPIGTIRYLTRNL 735

RESULT 8
 AAB50326
 ID AAB50326 standard; protein; 734 AA.

XX AC AAB50326;
 XX DT 09-MAR-2001 (first entry)
 XX DE Adeno-associated virus capsid protein sequence.
 XX KW Adeno-associated virus; AAV; capsid; virus binding inhibition;
 CC competitive inhibitor.

```

XX Masadenovirus.
OS Synthetic.
XX MO200073316-A2.
XX 07-DEC-2000.
XX 26-MAY-2000; 2000MO-US14466.
XX 28-MAY-1999; 99US-0321589.
XX (CELL-) CELL GENESYS INC.
XX
XX Patel S, McArthur J;
XX
XX WPI, 2001-061507/07.
XX
XX New polypeptide portion of a virus protein binding to an antibody
XX specific for the virus useful for inhibiting binding of the virus to a
XX cell or for binding host antibody to provide a transient tolerant or
XX non-responsive state.
XX
XX Disclosure; Fig 2; 33pp; English.
XX
XX The present sequence is given in a specification relating to
XX polypeptide portions of a virus protein or its derivative, that bind to
XX an antibody specific for the virus or inhibit binding of the virus to a
XX cell. The polypeptides are used in inhibiting the binding of viruses to
XX cells of a host. Oligopeptides that inhibit binding of viruses to
XX receptor can be used as competitive inhibitors to release bound virus
XX in an adsorption-type assay and if an antibody was used as an
XX immunoadsorbent, the oligopeptide could be used to elute bound virus
XX from a solid support to which virus antibody is immobilised. These
XX oligopeptides may further be used to bind to host antibody to provide a
XX transient tolerant or non-responsive state.
XX
XX Sequence 734 AA:
XX
XX Query Match 84.6%; Score 3376; DB 22; Length 734;
XX Best Local Similarity 83.0%; Pred. No. 1.7e-262;
XX Matches 611; Conservativity 51; Mismatches 72; Indels 2; Gaps 2;
XX
XX 1 MADGYLPDWLENDISEGIREMMDIKRGAPKPKANOQKODGGRGIVLPGYKYLGPRNGLD 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1 MADGYLPDWLENDISEGIREMMDIKRGAPKPKANOQKODGGRGIVLPGYKYLGPRNGLD 60
XX
XX 61 KGEPVNAADAALAEHDKAYDOOLKAGDNPYLRYNHADAEPQERLOEDTSFGNIGRAVFG 120
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 KGEPVNAADAALAEHDKAYDOOLKAGDNPYLRYNHADAEPQERLOEDTSFGNIGRAVFG 120
XX
XX 61 KGEPVNAADAALAEHDKAYDOOLKAGDNPYLRYNHADAEPQERLOEDTSFGNIGRAVFG 120
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 KGEPVNAADAALAEHDKAYDOOLKAGDNPYLRYNHADAEPQERLOEDTSFGNIGRAVFG 120
XX
XX 121 AKRRVLEPIGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTQOQPAKKRLNFGQDSE 180
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 121 AKRRVLEPIGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTQOQPAKKRLNFGQDSE 180
XX
XX 121 AKRRVLEPIGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTQOQPAKKRLNFGQDSE 180
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 121 AKRRVLEPIGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTQOQPAKKRLNFGQDSE 180
XX
XX 181 SVDPDPLGEPAPATPAVGTTPMAGSGGAPMANNBGAADVAGNAGWHDSTWMDGRVI 240
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 181 SVDPDPLGEPAPATPAVGTTPMAGSGGAPMANNBGAADVAGNAGWHDSTWMDGRVI 240
XX
XX 241 TTSRTWALPTYNHNLTKQISSASTGASNDNHFYGYSTPMGYDPRNFCHFPRIWQRL 300
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 241 TTSRTWALPTYNHNLTKQISSASTGASNDNHFYGYSTPMGYDPRNFCHFPRIWQRL 300
XX
XX 241 TTSRTWALPTYNHNLTKQISSASTGASNDNHFYGYSTPMGYDPRNFCHFPRIWQRL 300
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 241 TTSRTWALPTYNHNLTKQISSASTGASNDNHFYGYSTPMGYDPRNFCHFPRIWQRL 300
XX
XX 301 INNNWGRPRKLNPKLFNIOVKEVTNDGVTIANNLSTVQVPSDEYQLPYVLGSAHQ 360
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 301 INNNWGRPRKLNPKLFNIOVKEVTNDGVTIANNLSTVQVPSDEYQLPYVLGSAHQ 360
XX
XX 300 INNNWGRPRKLNPKLFNIOVKEVTNDGVTIANNLSTVQVPSDEYQLPYVLGSAHQ 359
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 300 INNNWGRPRKLNPKLFNIOVKEVTNDGVTIANNLSTVQVPSDEYQLPYVLGSAHQ 359
XX
XX 361 GCIPPPADVPMIPOYCYLTLNNGSQAVGSSFPYCLEFSPQMLRTNNNTFSTTEPVP 420
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 361 GCIPPPADVPMIPOYCYLTLNNGSQAVGSSFPYCLEFSPQMLRTNNNTFSTTEPVP 420
XX
XX 360 GCIPPPADVPMIPOYCYLTLNNGSQAVGSSFPYCLEFSPQMLRTNNNTFSTTEPVP 419
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 360 GCIPPPADVPMIPOYCYLTLNNGSQAVGSSFPYCLEFSPQMLRTNNNTFSTTEPVP 419
XX
XX 421 FHSSVYHQSGLDLNMLNPLIDQYLYLNRITONOGSAGNKKDLLFSRGSAPAGSVQPKWLP 480
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 421 FHSSVYHQSGLDLNMLNPLIDQYLYLNRITONOGSAGNKKDLLFSRGSAPAGSVQPKWLP 480
XX

```

```

DB 420 FHSSVYHQSGLDLNMLNPLIDQYLYLNRITNTSGTTGRLQFSQAGASIDRDSHMLP 479
XX
XX 481 GPCYRQORVSKTKTDNNNNNFMTGASKRYNLGRESIINPRGAMASHKDEDEFFPMSCV 540
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 481 GPCYRQORVSKTKTDNNNNNFMTGASKRYNLGRESIINPRGAMASHKDEDEFFPMSCV 540
XX
XX 480 GPCYRQORVSKTSADNNNSYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEFFPMSCV 539
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 480 GPCYRQORVSKTSADNNNSYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEFFPMSCV 539
XX
XX 541 MIFGKESAGASNTALDNWMITDEEIEKATNPVATERFGVAVNFQSSSTDPAVDYHMG 600
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 541 MIFGKESAGASNTALDNWMITDEEIEKATNPVATERFGVAVNFQSSSTDPAVDYHMG 600
XX
XX 540 LIFGKOGSEKTNVDIKMNTDEEIRTTN-VATERGYSVTNLQNGNQAATADVNTQG 598
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 540 LIFGKOGSEKTNVDIKMNTDEEIRTTN-VATERGYSVTNLQNGNQAATADVNTQG 598
XX
XX 601 ALPGWVMDRDVYLQGPIMAKIPIHTDGHFHPSPLMGSGFKLKPPIILIKNTFVPANPPA 660
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 601 ALPGWVMDRDVYLQGPIMAKIPIHTDGHFHPSPLMGSGFKLKPPIILIKNTFVPANPPA 660
XX
XX 661 EBSATKESAFITQYSTGVSVIEIEMELQENSGRNPPEVOYTSNAYKSNVDPVTNNGL 720
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 661 EBSATKESAFITQYSTGVSVIEIEMELQENSGRNPPEVOYTSNAYKSNVDPVTNNGL 720
XX
XX 659 TESAARFASFIQYSTGVSVIEIEMELQENSKRNMPETQYTSNAYKSNVDPVTNNGL 718
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 659 TESAARFASFIQYSTGVSVIEIEMELQENSKRNMPETQYTSNAYKSNVDPVTNNGL 718
XX
XX 721 YTEPRPIGTRVLRPL 736
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 721 YTEPRPIGTRVLRPL 736
XX
XX 719 YSEPRPIGTRVLRPL 734
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 719 YSEPRPIGTRVLRPL 734
XX
XX
XX RESULT 9
XX AAY71168
XX ID AAY71168 standard; protein; 599 AA.
XX
XX AC AAY71168;
XX
XX DT 08-SEP-2000 (first entry)
XX
XX DE Adeno-associated virus serotype 1 capsid protein VP2.
XX
XX KM Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX cap protein; recombinant viral vector; gene delivery; gene therapy;
XX vaccine; transgene; VP2.
XX
XX OS Adeno associated virus serotype 1.
XX
XX PN WO200028061-A2.
XX
XX PD 18-MAY-2000.
XX
XX PF 02-NOV-1999; 99MO-US25694.
XX
XX PR 05-NOV-1998; 98US-0107114.
XX
XX PA (UYPE-) UNIV PENNSYLVANIA.
XX
XX PI Wilson JM, Xiao W;
XX
XX WPI, 2000-376571/32.
XX
XX N-PSDB; AAD00772, AAD00778.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
XX
XX Claim 7; Page 93-95; 108pp; English.
XX
XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is an AAV-1 cap protein VP2 which is
XX useful in the production of recombinant viral vector for gene delivery.
XX

```

XX SQ Sequence 599 AA;
Query Match 81.5%; Score 3251; DB 21; Length 599;
Best Local Similarity 100.0%; Pred. No. 1,4e-252;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 TAPGKKRPVEOSPOBPSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPPTAPAA 197
DB 1 TAPGKKRPVEOSPOBPSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPPTAPAA 60
QY 198 VGPTTMAAGGAPADNNEGADGVGNASGNWCHDSYTLGDRVITTSRTTMAALPTNNHLY 257
DB 61 VGPTTMAAGGAPADNNEGADGVGNASGNWCHDSYTLGDRVITTSRTTMAALPTNNHLY 120
QY 258 KOISASTAGSNDNHYRGYSTPMGYPFDRNFRCHSPRDMQRLINNNGWFRPKRLNFKLF 317
DB 121 KOISASTAGSNDNHYRGYSTPMGYPFDRNFRCHSPRDMQRLINNNGWFRPKRLNFKLF 180
QY 318 NIQVKEVTTNDGVTTIANNLTSTVOVPSDSYOLPYVLGSAHQGLPPFPADVEMIPQY 377
DB 181 NIQVKEVTTNDGVTTIANNLTSTVOVPSDSYOLPYVLGSAHQGLPPFPADVEMIPQY 240
QY 378 YLTLLNGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSJLDRLLMP 437
DB 241 YLTLLNGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSJLDRLLMP 300
QY 438 LIDQYLYTANTQNGSAQKNDLLFSRGSPPAGMSVQPKMLPGPCYRQORVSKTKTDNN 497
DB 301 LIDQYLYTANTQNGSAQKNDLLFSRGSPPAGMSVQPKMLPGPCYRQORVSKTKTDNN 360
QY 498 NSNFTWTCASKYNLNGRESIINPGTAMASHKDDDEKFPFMSGWMIFGKESAGASTALDN 557
DB 361 NSNFTWTCASKYNLNGRESIINPGTAMASHKDDDEKFPFMSGWMIFGKESAGASTALDN 420
QY 558 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPAEDVHAMGALPGWMDRDVYLQGP 617
DB 421 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPAEDVHAMGALPGWMDRDVYLQGP 480
QY 618 IMALIPHTDGHFHSPLMGFGKLNKPPQILLIKTVPANPAPAESATKFASTFYQSTG 677
DB 481 IMALIPHTDGHFHSPLMGFGKLNKPPQILLIKTVPANPAPAESATKFASTFYQSTG 540
QY 678 QVSVEIEMELOKENSKRNPVEQYTSNAYAKSANDFTVDNGLYTEPRPIGRTYLRPL 736
DB 541 QVSVEIEMELOKENSKRNPVEQYTSNAYAKSANDFTVDNGLYTEPRPIGRTYLRPL 599
RESULT 10
AA71169
ID AAY71169 standard; Protein; 534 AA.
XX
AC AAY71169;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 capsid protein VP3.
XX
KM Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KM cap protein; recombinant viral vector; gene delivery; gene therapy;
KM vaccine; transgene; VP3.
XX
OS Adeno associated virus serotype 1.
XX
OS WO200028061-A2.
XX
PD 18-MAY-2000.
XX
PF 02-NOV-1999; 99WO-US25694.
XX
PR 05-NOV-1998; 98US-0107114.
XX
PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Wilson JM, Xiao W;
XX DR WPI; 2000-376571/32.
XX DR N-PSDB; AAD00772, AAD00779.
XX PT Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
PS Claim 7; Page 99-101; 108bp; English.
XX
XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterized by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is an AAV-1 cap protein VP3 which is
XX useful in the production of recombinant viral vector for gene delivery.
SQ Sequence 534 AA;
Query Match 72.9%; Score 2906; DB 21; Length 534;
Best Local Similarity 100.0%; Pred. No. 6,8e-225;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 MASGGAPADNNEGADGVGNASGNWCHDSYTLGDRVITTSRTTMAALPTNNHLYKOISS 262
DB 1 MASGGAPADNNEGADGVGNASGNWCHDSYTLGDRVITTSRTTMAALPTNNHLYKOISS 60
QY 263 ASTGASNDNHYRGYSTPMGYPFDRNFRCHSPRDMQRLINNNGWFRPKRLNFKLFNIQVK 322
DB 61 ASTGASNDNHYRGYSTPMGYPFDRNFRCHSPRDMQRLINNNGWFRPKRLNFKLFNIQVK 120
QY 323 EVTTNDGVTTIANNLTSTVOVPSDSYOLPYVLGSAHQGLPPFPADVEMIPQYGYLTNN 382
DB 121 EVTTNDGVTTIANNLTSTVOVPSDSYOLPYVLGSAHQGLPPFPADVEMIPQYGYLTNN 180
QY 383 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSJLDRLLMPIDQY 442
DB 181 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSJLDRLLMPIDQY 240
QY 443 LYYLANTQNGSAQKNDLLFSRGSPPAGMSVQPKMLPGPCYRQORVSKTKTDNNNSFT 502
DB 241 LYYLANTQNGSAQKNDLLFSRGSPPAGMSVQPKMLPGPCYRQORVSKTKTDNNNSFT 300
QY 503 WTGASKYNLNGRESIINPGTAMASHKDDDEKFPFMSGWMIFGKESAGASTALDNWITD 562
DB 301 WTGASKYNLNGRESIINPGTAMASHKDDDEKFPFMSGWMIFGKESAGASTALDNWITD 360
QY 563 EBEIKATNPVATERFGTVAVNFQSSSTDPAEDVHAMGALPGWMDRDVYLQGPIMAKI 622
DB 361 EBEIKATNPVATERFGTVAVNFQSSSTDPAEDVHAMGALPGWMDRDVYLQGPIMAKI 420
QY 623 PHTDGHFHSPLMGFGKLNKPPQILLIKTVPANPAPAESATKFASTFYQSTGQSV 662
DB 421 PHTDGHFHSPLMGFGKLNKPPQILLIKTVPANPAPAESATKFASTFYQSTGQSV 480
QY 663 IEMELOKENSKRNPVEQYTSNAYAKSANDFTVDNGLYTEPRPIGRTYLRPL 736
DB 481 IEMELOKENSKRNPVEQYTSNAYAKSANDFTVDNGLYTEPRPIGRTYLRPL 534
RESULT 11
AAG65793
ID AAG65793 standard; Protein; 598 AA.

AC AAG65793;
 XX 11-FEB-2002 (first entry)
 DT XX
 DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.
 XX
 KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
 KM inverted terminal repeat; nootropic; neuroprotective; anti-neurotic; ITR;
 KM antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
 KM major coat protein; AAV-2; VP2.
 XX
 OS Adeno-associated virus 2.
 XX
 FH Key
 FT Misc-difference 1
 FT /note= "encoded by ACG"
 XX
 PN WO200168888-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 13-MAR-2001; 2001WO-US07927.
 XX
 PR 14-MAR-2000; 2000US-189110P.
 XX
 PA (NEUR-) NEUROLOGIX INC.
 XX
 PI Xiao W, During MJ;
 XX
 DR WPI; 2001-596912/67.
 DR N-PSDB; AAI66974.
 XX
 PT Recombinant viral vector useful in improving gene therapy in a subject,
 PT and for increasing efficiency of entry into a cell, comprises a
 PT chimeric capsid having one non-native amino acid sequence and a desired
 PT transgene
 PT
 PS Disclosure; Page, 51; 53pp; English.
 PS
 XX The invention provides a recombinant viral vector (RVV) comprising a
 CC chimeric capsid (I) having at least one non-native amino acid sequence,
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),
 CC or their combination, and a transgene flanked 5' and 3' by inverted
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their
 CC combination. The RVV is useful for improving gene therapy in a subject
 CC with a disorder, and for increasing the efficiency of entry into a cell,
 CC which involves producing (I) encapsulating a viral vector, and contacting
 CC a cell with RVV having (I) such that (I) binds to an attachment site on
 CC the cell surface and permits the vector to enter the cell efficiently.
 CC A pharmaceutical composition comprising RVV with (I) containing a
 CC transgene sequence associated with a disease or a disorder such that
 CC expression of the transgene would result in amelioration of the disease
 CC or disorder such as inherited neurological and metabolic diseases e.g.,
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,
 CC retinoblastoma and various types of neoplastic cells which include,
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,
 CC sarcomas, leukemias and lymphoma. The present sequence represents the
 CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences
 CC are used in the construction of a chimeric vector.
 CC
 SO Sequence 598 AA;

Query Match 69.2%; Score 2759.5; DB 22; Length 598;
 Best Local Similarity 82.8%; Pred. No. 4.9e-213;
 Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGRKRVESPOEPDSSSGTGTGQOPAKRLNFGTGTGSESVPPDPQIPGEPATPAV 198
 DB 2 APGRKRVESVPEPSSSGTGTGQOPAKRLNFGTGTGSESVPPDPQIPGEPATPAV 61

QY 199 GPTMASGGGAPMADNNEGADGVGNASGNWCHDSTVGLDRVITTTSTRTALPTYNHLYK 258
 DB 62 GTMTATGSGAPMADNNEGADGVGNASGNWCHDSTVGLDRVITTTSTRTALPTYNHLYK 121
 QY 259 QISSASTGASNDNHYFGYSTPWGYPFDRPHCHFSRPMQRLINNMGFRPKLNFLEN 318
 DB 122 QISSQS-GASNDNHYFGYSTPWGYPFDRPHCHFSRPMQRLINNMGFRPKLNFLEN 180
 QY 319 IQVKEVTTNDGVTIANNLSTVQVSDSEYQLPYLGSAGQCLPPPADVMTIQYGY 378
 DB 181 IQVKEVTTNDGVTIANNLSTVQVSDSEYQLPYLGSAGQCLPPPADVMTIQYGY 240
 QY 379 LITNNGSOAVGRSFPCLXFPSPQMLRTGNPFSTFPEEPHSSYASQSIDRLMNP 438
 DB 241 LITNNGSOAVGRSFPCLXFPSPQMLRTGNPFSTFPEEPHSSYASQSIDRLMNP 300
 QY 439 IDQVLYYLNRTONQSSAOKDILFSGSPAGMSVOPKWLPGPCYRQORVSKTKTDNNN 498
 DB 301 IDQVLYYLNRTONQSSAOKDILFSGSPAGMSVOPKWLPGPCYRQORVSKTKTDNNN 360
 QY 499 SNETWTGASKYLNNGRESIINPCTAMASHKDEDEKFPMSGVMIKESAGASNTALDNY 558
 DB 361 SEXSWTGATCYHLNGRDSLVPGPAMASHKDEDEKFPMSGVMIKESAGASNTALDNY 420
 QY 559 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVAMGALPGMWMODRVPYLOGP 618
 DB 421 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVAMGALPGMWMODRVPYLOGP 480
 QY 619 MAKIPHTDGHFHPSPMLMGFGKLNPPQILIKNTVPANPPAFSAKFSPTQYSTGQ 678
 DB 481 MAKIPHTDGHFHPSPMLMGFGKLNPPQILIKNTVPANPPAFSAKFSPTQYSTGQ 540
 QY 679 VSVEIEMWELQENSKRNREVOYTSNYAKSANDVFDVNNGLTERPPIGRTYLTPL 736
 DB 541 VSVEIEMWELQENSKRNREVOYTSNYAKSANDVFDVNNGLTERPPIGRTYLTPL 598

RESULT 12

AAAS1509
 ID AAM51509 standard; protein; 598 AA.

AC AAM51509;

DT 02-JUN-2002 (first entry)

DE Adeno-associated virus VP2 capsid protein.

KM Adeno-associated virus; AAV; VP2; capsid; virus-like particle;

KM nuclear localisation signal; VP3.

OS Adeno associated virus.

PN JP2001169777-A.

PD 26-JUN-2001.

PF 30-JUL-1999; 99JP-0249140.

PR 30-JUL-1999; 99JP-0249140.

PA (HAND/) HANDA H.

DR WPI; 2001-599854/68.

PT New virus-like particles from VP3 capsid protein of adeno-associated

PT virus, comprise a peptide containing a nucleus-shifting signal

PS Claim 1; Page 14-16; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a
 CC virus-like particle-forming protein, and to a peptide containing a
 CC nuclear-shifting signal at its N-terminus that can form virus-like

CC particles by shifting to the nucleus of the animal cell in which it
CC is expressed. The method is used for forming virus-like particles
CC from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 598 AA;

Query Match 69.2%; Score 2759.5; DB 22; Length 598;

Best Local Similarity 82.8%; Pred. No. 4.9e-213; Indels 1; Gaps 1;
Matches 495; Conservative 43; Mismatches 59;

QY 139 APGKKRPVEOSPOEPPSSSGIGKTGQOQPAKRLNFGOTGSESVDDPQEGEPATPAV 198
DB 2 APGKKRPVEHSPVPDSSSGTGKAGQOQPARKRLNFGOTGADSVDDPQEGEPAPASGL 61
QY 199 GPTTMAAGGAPMADNNEGADGVCNAGNHHCDSTWLGDRITTTSTWALPTNNHLYK 258
DB 62 GTNMAAGGAPMADNNEGADGVCNAGNHHCDSTWLGDRITTTSTWALPTNNHLYK 121
QY 259 QISSASTGASNDNHYFGYSTPMGYFDENRPFCHFSPPDMORLINNNNGFPRKRLNFKL 318
DB 122 QISSQS-GASNDNHYFGYSTPMGYFDENRPFCHFSPPDMORLINNNNGFPRKRLNFKL 180
QY 319 IQVEVTNDGVTIANNLSTVOVFSDEYQLPVLSAHQGLPPPADVEMIPOYGY 378
DB 181 IQVEVTQNDGVTIANNLSTVOVFTDSEYQLPVLSAHQGLPPPADVEMVPOYGY 240
QY 379 LTLNNGSOAVGRSSFYCLEYFPSSOMLRTGNNFTSYFEEYFHSVAHQSILRLNPL 438
DB 241 LTLNNGSOAVGRSSFYCLEYFPSSOMLRTGNNFTSYFEEYFHSVAHQSILRLNPL 300
QY 439 IDQYLYLNRTONSGSAQNKDLFSRGSFAGMSVQPKMLFCPCYQOQVSKTKTDNN 498
DB 301 IDQYLYLNRTONSGSTTQSRLOFSQAGADINDQSNMFLFCYQOQVSKTKSADNN 360
QY 499 SNFTWTGASKTNLNGRESINPGTAMASHKODEKFFPMGCVMTFGESAGASTALDNV 558
DB 361 SEYMTGATKTHLGRDSTLVNPGPAMASHKODEKFFPMGCVMTFGESAGASTALDNV 420
QY 559 MITDEERKATNPATBERFGTVAVNFOSSSTDPATGVDHANGALPGWVODRDVYLGP 618
DB 421 MITDEERKATNPATBERFGTVAVNFOSSSTDPATGVDHANGALPGWVODRDVYLGP 480
QY 619 WAKIPTHGHPPLMGFGFLKKNPPQILIKNTVPANPAPAFSATKFAFIQYSTGQ 678
DB 481 WAKIPTHGHPPLMGFGFLKKNPPQILIKNTVPANPAPAFSATKFAFIQYSTGQ 540
QY 679 VSVIEIEMELQENSKRNPNVEQYTSNTAKSANVDFTVDNNGLYTEPRPIGRYITRPL 736
DB 541 VSVIEIEMELQENSKRNPNVEQYTSNTAKSANVDFTVDNNGLYTEPRPIGRYITRPL 598

RESULT 13

AAW46308 ID AAW46308 standard; Protein; 734 AA.

XX AC AAW46308;

XX DT 28-AUG-1998 (first entry)

XX DE AAW4 VP1 capsid protein.

XX KM AAV; AAV-4; VP1; capsid protein; vector; gene transfer;

XX KW gene delivery; cancer; gene therapy.

XX OS Adeno associated virus 4.

XX FH Key Location/Qualifiers

FT Misc-difference 208 /note= "encoded by GGC"

FT Misc-difference 250 /note= "encoded by CTN"

XX FT
XX PN W09811244-A2.

XX 19-MAR-1998.

XX 11-SEP-1997; 97MO-US16266.

XX 11-SEP-1996; 96US-0025934.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chiorini JA, Kotin RM, Safer B;

XX WPI; 1998-207403/18.

XX N-PSDB; AAV21648 AND AAV21650.

XX Adeno-associated virus 4 based vectors - used to transduce erythroid
PT progenitor cells for treatment of e.g. cancer and other genetic
PT diseases

XX Claim 34; Page 54-55; 80pp; English.

XX This polypeptide comprises the adeno associated virus 4 (AAV4)
XX VP1 capsid protein that is encoded by open reading frame 2 (see
XX AAV21650) of the AAV4 genome (see also AAV21648). The invention also
XX provides AAV4 VP2 (see AAW46313) and VP3 (see AAW46314), also encoded
XX by the AAV4 genome (see also AAV21657-58). The invention provides
XX recombinant vectors and viral particles based on AAV4 that may be
XX useful for transducing erythroid progenitor cells for the treatment
XX of e.g. cancer and genetic diseases which can be corrected by bone
XX marrow transplants using matched donors. A claimed method of
XX delivering a nucleic acid to a subject (including a subject with
XX antibodies to AAV2) comprises administering to a cell from the
XX subject an AAV4 particle comprising the nucleic acid inserted
XX between a pair of AAV inverted terminal repeats (see AAV21651 and
XX AAV21659), and returning the cell to the subject.

XX Sequence 734 AA;

Query Match 62.3%; Score 2486.5; DB 19; Length 734;

Best Local Similarity 63.5%; Pred. No. 6e-191; Indels 31; Gaps 10;
Matches 475; Conservative 81; Mismatches 161;

QY 4 DGYLPWLEDNLSSEGIREWMDLKPAPKPKRANOQKDDGRLVLPQYKTYGPNGLDKGE 63
DB 3 DGYLPWLEDNLSSEGIREWMDLKPAPKPKRANOQKDDGRLVLPQYKTYGPNGLDKGE 62
QY 64 PVNAADAAALAEHKAADQOQKAGDNPLYLRNHADAFOEJLOEDTSFGGNLGRAVQAKK 123
DB 63 PVNAADAAALAEHKAADQOQKAGDNPLYLRNHADAFOEJLOEDTSFGGNLGRAVQAKK 122
QY 124 RYLEPLGLVEEAGKTPAGKKRPVEOSPOEPPSSSGIGKTGQOQPAKRLNF-GOTGSESV 182
DB 123 RYLEPLGLVEEAGKTPAGKKRPVEOSPOEPPSSSGIGKTGQOQPAKRLNFVEEDTGAQDGE 182
QY 183 PDPQEGEPATPAVAVGPTTMAAGGAPMADNNEGADGVCNAGNHHCDSTWLGDRYITT 242
DB 183 PEGSTSG-----AMSDDSEMRRAAGAAVEGGGADGVNAGDMHCDSTWSEGVYTTT 236
QY 243 STRTWMALPTNNHLYKQOISSASTGASNDNHYFGYSTPMGYFDENRPFCHFSPPDMORLIN 302
DB 237 STRTWMALPTNNHLYKRLGE---SLQSNTRYNFSFPMGYFDENRPFCHFSPPDMORLIN 292
QY 303 NNMGFRPKRLNFKLFINIQVEVTNDGVTIANNLSTVOVFSDEYQLPVLSAHQGL 362
DB 293 NNMGFRPKRMRYLFINIQVEVTTSNGETTIVANNLSTVOIFADSSYELPYWDAQEGS 352
QY 363 LPPFPADVEMIPOYGY---LTLNNGSOAVGRSSFYCLEYFPSSOMLRTGNNFTSYFEEY 419
DB 353 LPPFPADVEMVPOYGCGLVGTGNTSQOQDRNMFYCLEYFPSSOMLRTGNNFETYSFEKV 412
QY 420 PHSSTVAHSQSLRLNPNPLIDQYLYLNRTON---QSGSAQNKDLFSRGSFAGMSVOP 475
DB 413 PHSMAVHQSILRLNPNPLIDQYLYLNRTON---QSGSAQNKDLFSRGSFAGMSVOP 469

QY 476 KMLPGCFYRQORVSKTKTDNNNSNFTWTGAS---KY-----NINGRESIINPGTAMASHK 528
 DB 470 KMLPGESIRIQGGSKTA--NQNFKIATGSDSLIKETHTSTLDGRKMSALTPEPMATAG 527
 QY 529 DDEKDFPMGVMIFGKESAGSNTALDNWITDEEIKATNPVATERFSTVAVNFGSSS 588
 DB 528 PADS-K-FSNSQLIFAGPKQNGNTATVGTGLIFTESEELATNATDTDMGNLPGDQSN 586
 QY 589 TDPATGVHMGALPGVNMDDRDVYLOGPIWAKIPTHGDFHPSPLMGFGGLKNPPQIL 648
 DB 587 NLPTVDRLTALGAVPGVWQNRDLYYGPIWAKIPTHGDFHPSPLMGFGGLKNPPQIL 646
 QY 649 IKNTPVANPAESATKFASTFOYSTGVSEIEMELOKENSKRWNPVQYTSNYSKS 708
 DB 647 IKNTPVANPAETPSSPTVNSFTIOYSTGVSDIOWEIOKERSKRWNPVQYTSNYSQ 706
 QY 709 ANVDFTVNNGLYTPRPITGRVLTPL 736
 DB 707 NSLIMAPDAAGKYTEPRATIGRTYLTHTL 734

RESULT 14
 AAG65794
 ID AAG65794 standard; Protein; 533 AA.
 XX
 AC AAG65794;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Adeno-associated virus 2 (AAV-2) major coat protein VP3.
 XX
 KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
 KM inverted terminal repeat; nonretroviral; neuroprotective; anti-neurotic;
 KM antitumor; anticancer; gene therapy; adeno-associated virus; AAV;
 KM major coat protein; AAV-2; VP3.
 XX
 OS Adeno-associated virus 2.
 XX
 PN WO200168888-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 13-MAR-2001; 2001WO-US07927.
 XX
 PR 14-MAR-2000; 2000US-189110P.
 XX
 PA (NEUR-) NEUROLOGIX INC.
 PI Xiao W, During MJ;
 XX
 DR WPI; 2001-596912/67.
 DR N-PSDB; AAI66974.
 XX
 XX
 PT Recombinant viral vector useful in improving gene therapy in a subject,
 PT and for increasing efficiency of entry into a cell, comprises a
 PT chimeric capsid having one non-native amino acid sequence and a desired
 PT transgene -
 XX
 PS Disclosure; Page 51; 53pp; English.
 XX
 CC The invention provides a recombinant viral vector (RVV) comprising a
 CC chimeric capsid (I) having at least one non-native amino acid sequence,
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),
 CC or their combination, and a transgene flanked 5' and 3' by inverted
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their
 CC combination. The RVV is useful for improving gene therapy in a subject
 CC with a disorder, and for increasing the efficiency of entry into a cell,
 CC which involves producing (I) encapsulating a viral vector, and contacting
 CC a cell with RVV having (I) such that (I) binds to an attachment site on
 CC the cell surface and permits the vector to enter the cell efficiently.
 CC A pharmaceutical composition comprising RVV with (I) containing a
 CC transgene sequence associated with a disease or a disorder such that
 CC expression of the transgene would result in amelioration of the disease

CC or disorder such as inherited neurological and metabolic diseases e.g.
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,
 CC retinoblastoma and various types of neoplastic cells which include
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,
 CC sarcomas, leukemias and lymphoma. The present sequence represents the
 CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences
 CC are used in the construction of a chimeric vector.
 XX
 SQ Sequence 533 AA;
 CC
 CC Query Match 62.2%; Score 2481.5; DB 22; Length 533;
 CC Best Local Similarity 83.3%; Pred. No. 9,4e-191;
 CC Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;
 QY 203 MASGGAPADNNEGADGVNAGSNWCHDSTWLGDRVITTSRTTMAIPYNNHLYKQISS 262
 DB 1 MATGSGAPADNNEGADGVNAGSNWCHDSTWLGDRVITTSRTTMAIPYNNHLYKQISS 60
 QY 263 ASTGASNDNHYFGYSTPWCYFDFNRFHCHSPRDMQRLINNMGFRPKLNFYLIQYK 322
 DB 61 QS-GASNDNHYFGYSTPWCYFDFNRFHCHSPRDMQRLINNMGFRPKLNFYLIQYK 119
 QY 323 EYTTNDGVTTIANNLSTVOVPSDSEYOLPYVIGSAHOCCLPRPADVEMIPQYGLTIN 382
 DB 120 EYTGNDGVTTIANNLSTVOVPSDSEYOLPYVIGSAHOCCLPRPADVEMIPQYGLTIN 179
 QY 383 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSQSLDRMLNPLIDY 442
 DB 180 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEEVPHSSYAHQSQSLDRMLNPLIDY 239
 QY 443 LYLINFTONQSGSAQKDLIFSRGSPAGMSVOPKMLPGPCYRQGRVSKTKTDNNNSNFT 502
 DB 240 LYLISRTNTPSGTTSRQLQFSQAGASDIRDSRNLPGPCYRQGRVSKTSADNNSEYS 299
 QY 503 WTGASKYNNNGRESIINPGTAMASHKDDKDFPMGVMIFGKESAGSNTALDNWITD 562
 DB 300 WTGATKYHLNGRSLVNPVAPAMASHKDDKDFPMGVMIFGKESAGSNTALDNWITD 359
 QY 563 EEEIKATNPVATERFSTVAVNFGSSSTDPATGVHMGALPGVNMDDRDVYLOGPIWAKI 622
 DB 360 EEEIRTTNPVATERFSTVAVNFGSSSTDPATGVHMGALPGVNMDDRDVYLOGPIWAKI 419
 QY 623 PHTDGHFSPPLMGFGGLKNPPQILIKNTPVANPAESATKFASTFOYSTGVSE 682
 DB 420 PHTDGHFSPPLMGFGGLKNPPQILIKNTPVANPAESATKFASTFOYSTGVSE 479
 QY 683 IEMELOKENSKRWNPVQYTSNYSKASANDFTVDNNGLYTPRPITGRVLTPL 736
 DB 480 IEMELOKENSKRWNPVQYTSNYSKASANDFTVDNNGLYTPRPITGRVLTPL 533

RESULT 15
 AAM51510
 ID AAM51510 standard; protein; 533 AA.
 XX
 AC AAM51510;
 XX
 DT 02-JAN-2002 (first entry)
 XX
 DE Adeno-associated virus VP3 capsid protein.
 XX
 KM Adeno-associated virus; AAV; capsid; virus-like particle;
 KM nuclear localisation signal; VP3.
 XX
 OS Adeno associated virus.
 XX
 PN JP2001169777-A.
 XX
 PD 26-JUN-2001.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 18.1145 Seconds
(without alignments)
1195.466 Million cell updates/sec

Title: US-09-807-802A-13

Sequence: 1 MADGYLPMLEDNLSGIR.....NNGLYTEPRPIGTRVLRPL 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A COMB .pep: *
2: /cgn2_6/ptodata/1/1aa/5B COMB .pep: *
3: /cgn2_6/ptodata/1/1aa/6A COMB .pep: *
4: /cgn2_6/ptodata/1/1aa/6B COMB .pep: *
5: /cgn2_6/ptodata/1/1aa/PTUS COMB .pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1 .pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2486.5	62.3	734	US-09-532-594B-4	Sequence 4, Appl1
2	1630.5	45.9	598	US-09-532-594B-16	Sequence 16, Appl1
3	1690.5	42.4	544	US-09-532-594B-18	Sequence 18, Appl1
4	479.5	12.0	543	US-08-856-841-22	Sequence 22, Appl1
5	430.5	10.8	500	US-08-856-841-16	Sequence 16, Appl1
6	430.5	10.8	501	US-08-856-841-18	Sequence 18, Appl1
7	429.5	10.8	456	US-08-856-841-19	Sequence 19, Appl1
8	326	8.2	415	US-08-856-841-20	Sequence 20, Appl1
9	318	8.0	395	US-08-856-841-13	Sequence 13, Appl1
10	310	7.8	264	US-08-856-841-14	Sequence 14, Appl1
11	303	7.6	398	US-08-856-841-21	Sequence 21, Appl1
12	284	7.1	387	US-08-856-841-17	Sequence 17, Appl1
13	216.5	5.4	579	US-09-022-949-2	Sequence 2, Appl1
14	192	4.8	584	US-09-022-949-2	Sequence 2, Appl1
15	142	3.6	210	US-08-856-841-9	Sequence 9, Appl1
16	142	3.6	227	US-08-856-841-15	Sequence 15, Appl1
17	142	3.6	250	US-08-856-841-12	Sequence 12, Appl1
18	124	3.1	3060	US-08-487-826B-14	Sequence 14, Appl1
19	119.5	3.0	1394	US-08-296-791-2	Sequence 2, Appl1
20	119.5	3.0	1394	PCT-US95-10661A-2	Sequence 2, Appl1
21	114.5	2.9	434	US-08-710-249-4	Sequence 4, Appl1
22	114.5	2.9	434	US-09-230-157A-4	Sequence 4, Appl1
23	113.5	2.8	655	US-08-469-202-27	Sequence 27, Appl1
24	113.5	2.8	655	US-08-484-434C-34	Sequence 34, Appl1
25	112.5	2.8	1096	US-09-415-946-3	Sequence 3, Appl1
26	110.5	2.8	624	US-08-947-965-78	Sequence 78, Appl1
27	110.5	2.8	655	US-08-469-202-28	Sequence 28, Appl1

28	110.5	2.8	655	US-08-484-434C-35	Sequence 35, Appl1
29	110.5	2.8	1751	US-09-136-574A-44	Sequence 44, Appl1
30	109.5	2.7	1651	US-09-540-245A-18	Sequence 18, Appl1
31	109.5	2.7	1848	US-08-296-791-6	Sequence 6, Appl1
32	109.5	2.7	1848	PCT-US95-10661A-6	Sequence 6, Appl1
33	109	2.7	1848	US-08-485-568A-4	Sequence 4, Appl1
34	109	2.7	1186	US-08-357-698-6	Sequence 6, Appl1
35	109	2.7	1186	US-08-590-554A-4	Sequence 4, Appl1
36	109	2.7	1186	US-09-184-223-4	Sequence 4, Appl1
37	109	2.7	1186	PCT-US93-12682-6	Sequence 6, Appl1
38	109	2.7	1545	US-08-296-791-4	Sequence 4, Appl1
39	109	2.7	1545	PCT-US95-10661A-4	Sequence 4, Appl1
40	107.5	2.7	331	US-08-793-426A-2	Sequence 2, Appl1
41	107.5	2.7	331	US-08-294-565-2	Sequence 2, Appl1
42	106.5	2.7	331	US-08-136-993-1	Sequence 1, Appl1
43	106.5	2.7	331	US-09-109-063-1	Sequence 1, Appl1
44	106.5	2.7	331	US-08-793-426A-3	Sequence 3, Appl1
45	106.5	2.7	331	US-09-294-565-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-532-594B-4

Sequence 4, Application US/09532594B

Patent No. 6468524

GENERAL INFORMATION:
APPLICANT: Chortini, John A.

APPLICANT: Kotlin, Robert M.

APPLICANT: Safer, Brian

APPLICANT: Davidson, Beverly

TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF

FILE REFERENCE: 14014.025202

CURRENT APPLICATION NUMBER: US/09/532,594B

CURRENT FILING DATE: 2000-03-22

NUMBER OF SEQ ID NOS: 22

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 734

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence; No. 6468524 =

NAME/KEY: misc:feature

OTHER INFORMATION: AAV4 capsid protein VPI

US-09-532-594B-4

Query Match

Best Local Similarity 63.3%; Score 2486.5; DB 4; Length 734;

Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

Query

4 DGYLPMLEDNLSGIRWMDLKPAPKPKANQKODDGLVLPYKYLGPFGDLKGE 63

3 DGYLPMLEDNLSGIRWMDLKPAPKPKANQKODDGLVLPYKYLGPFGDLKGE 62

64 PVNAADAALEHDKAVDQDLKAGDNPYLRYNHADEFFERLQEDTSFGGNLGRAVFOAKK 123

63 PVNAADAALEHDKAVDQDLKAGDNPYLRYNHADEFFERLQEDTSFGGNLGRAVFOAKK 122

124 RYLEPGLVEEGAKTPAGKKRPVEQSPORPDSSGIGTKGQOPAKKRLNF-GQTGSESV 182

123 RYLEPGLVEEGAKTPAGKKRPVEQSPORPDSSGIGTKGQOPAKKRLNF-GQTGSESV 182

183 PPGPLGEPPTAPPAVGTPTMAAGGAPADNNEGADVGNASGNNHCDSTWLGDRVIT 242

182 PPGPLGEPPTAPPAVGTPTMAAGGAPADNNEGADVGNASGNNHCDSTWLGDRVIT 242

243 STRTALLPYNNHLTKOISASTGASNDNHYGYSTPWGFPDPRNHCFFSPDWORLIN 302

237 STRTALLPYNNHLTKOISASTGASNDNHYGYSTPWGFPDPRNHCFFSPDWORLIN 302

Qy	303	NNNGFRKRLNFTLENIQVEVETINDGVTTIANNTLSITVOVSEDEYOLPVLGSAHQG	362
Db	293	NNMGSRKAKRVKIPNIQVEVYTSNGEFTVANLSTVOIFADBSYTELPVYMAQBGCS	352
Qy	363	LPPPPADVEMIPOYGY--LTLNNGSQAVGRSSFYCLEYFPGOMLRGNNFTFSYTEEV	419
Db	353	LPPPPNDVFWVPQYGCGLVGTGNTSQOQTRDNAFCYLEFPGOMLRGNNFEIYSEEV	412
Qy	420	PFFSSVYHSGSLRLRNPLIDQVLYLYLNTQN---QSGSANQMDLFRSRSRAGMSVOP	475
Db	413	PFFSHMAHSGSLRLRNPLIDQVLMELQSTGTTLANAGTATN---FTKLRPNFNFKF	465
Qy	476	KNNLPGECHYQQRFSKTKTIDNNNSNFTWGA---KY-----NLNGRSIINPTGAMASHK	528
Db	470	KNNLPGSIHQQGFSKTA--NQNYKIRAGSGSOLIKYETHSLDGRMGALIRPGPMATAG	527
Qy	529	DDEDKFFPMGCVMIFGKESAGASNTALDVMNMTDDEBIKATNPVATERPGTVAVNFOSS	588
Db	528	PAQSK-FSNSQLFPAGKQNGNATVPGLIFTESEELAAINTADTIDMGMLPGQDOSNS	586
Qy	589	TDPAITGVHMGALPGWVMDRVRVYQOGFIYMKITHTGHPHPSPLMGFGELKNRPQIL	648
Db	587	NLPVYDRLTLGAVPGKVMQNRDIYQSGFIYMKIHTDGHFRPSLIGSGFGELKNRPQIF	646
Qy	649	IKQTPVAPNPAEFSATKFPASFIYQYSTQGVSEVIELEWLOKENSGRNPNVEQVYTSNYSKS	708
Db	647	IKQTPVAPNPAITFSSIPVNSFITQYSTQGVSVQIDWEIQERSKRNPNVEQVTSNYGQ	706
Qy	709	ANVDFTVNDNGLTBERPITGRILTRPL	736
Db	707	NSLIIMAPDAAGKYTEPRAIGTRILTHHL	734

```

RESULT 2
; US-09-532-594B-16
; Sequence 16, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorin, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP2
; US-09-532-594B-16

```

	Query Match	45.9%	Score 1830.5;	DB 4;	Length 598;
	Best Local Similarity	57.8%;	Pred. No.6.3e-152;		
	Matches	355;	Conservative	74;	Mismatches 154; Indels 31; Gaps 10;
QY	138 TAPGKKRIVEQSPOEDSSSGIGTKGGQPAAKRLNF-GGTGHSSEVPPDPLGEPPATPA	196			
	: : : :				
Db	1 TAPGKKRRLPIESPQQPDSSTGI GKKGKQPAKKKLVEFDETAGDGDPPEGSTG-----A	54			
QY	197 AVGPRTMASGGGAAPADNNEGADGVGNASGNMHCOSTLGDIVITTSRTMALPTYYNHL	256			
	: : :				
Db	55 MSDDSEMFAAGAAVEBGGADGVGNASGDHCHCOSTSBEHVTTTSTRTWLPYYNHL	114			
	: : :				
QY	257 YKOISSASTGASNDNHFGYSIPWGYPPDNRRHFCHFSPPDWQLINNMMGFPRPKXNLFXL	316			
	: : :				

```

Db 115 YKJLGE-----SLQSNYNNGFSTPMGCFPFNFPHCHFSRBDQRLINNMCKRRKARVKI 170
Qy 317 FNIOVKEVYTNNGVTIANNTLSTVOVSDSEYQLPYVLGSAHOGCLPPEPAVFNIPQY 376
Db 171 FNIOVKEVTTSGEYTVANNLTSTVOYIADSSYELPYVMDGQGSLLPPEPNDFVMPQY 230
Qy 377 GY---LTLNNGSAOAVRS6FYCLEVFP6SOMRTGNPFSTFEEVPEFHSYAHOSLDR 433
Db 231 GYCGVLGTGMSQQQDDBRANAFYLEVFP6SOMRTGNPFETITSFKEVPFHSYAHOSLDR 290
Qy 434 LAMPILIDOTLYLYLNTQN---QSSAONKULLFSRGSBAGMSYQPKWMLPGFCYRQRY 489
Db 291 LAMPILIDYOYLMGLOSTTGGITLINAAGTATN---ETKLRPTNFSPFKKNWLP6SIRKQGF 347
Qy 490 SKTKTNNNSNFTWTCAS---KY-----NLNGRESTINPGTAMASHKODEKPFMSGVM 542
Db 348 SKTR--NOMYKPIRAGSDBSLKYEHTSLDGRMSALTPGPRMARA6ADSK-FSNSQLIF 404
Qy 543 FKESAGASNTALDNVMTDEEELKATNPVATERPVTAVNVFQSSSTDPAITGVHANGAL 602
Db 405 AGCRKONGNATVPGLTIFTSBEELALATATDTDMWGNLPGDDOSNSMLPTVDRJLALGAV 464
Qy 603 PGWVMDDBDVYLOGPIMAKIPHTDGHFSPSPLMG6FGKJNPPOILLKNTPVANPAPAEF 662
Db 465 PGWVMDORIDYIYQGPIMAKIPHTDGHFSPSPJIG6FGKJNPPOILFKNTPVANPAPITF 524
Qy 663 SATKFASTFOYSTQGVSVLEIMELQKENSKRANPEVOYITSYAKSANDVFTVDNNGLYT 722
Db 525 SSTPVNSFTYQSTQGVSVQIDMEIQKERSKRANPEVOFTSYNQOQSLMLAPDAAGKYT 584
Qy 723 EPRPIGTRVLYTRPL 736
Db 585 EPRPIGTRVLYTHHL 598

```

```

RESULT 3
US-09-532-594B-18
; Sequence 18, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotlin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAIV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAIV4 capsid protein VP3
; US-09-532-594B-18

```

[illegible]

QY 324 VTNDGVTIANLNTSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMIPKGY---LT 380
 DB 124 VTTNSNGTIANLNTSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMIPKGYCGLVT 183
 QY 381 LNNQSAVGSSEYCYLPPSPQMLRTGNNFTSYTPEEVPFSSYAHOSGLDRLMLDID 440
 DB 184 GNTSQOQTDHNAFYCLTFPSQMLRTGNNFTSYTPEEVPFSSYAHOSGLDRLMLDID 243
 QY 441 QYLYLNRQ---QGSQONDLFSRGSFAGMSQVQPKMLPGPCYRQORVSKXTDN 456
 DB 244 QYLMGLOSTYTTGLTANGLATN---PTKLPINFSNPKKNMLPGPSIKQGSFKA--N 298
 QY 497 NNSNFTWTGAS---KY---NLNGRESIINPGTAMASHKDEDEFFPMGSMIFGKESAG 549
 DB 299 QNKIPATGSDSLIKYTHSTLDORWALTPGPMAFAGADSK--FENSQILFAPGQNG 357
 QY 550 ASNTALDNNITDDEEIKATNPVATERFGVAVNFQSSSTDPTAGDVHANGALPGMYWOD 609
 DB 358 NTATVPGLTLFTSEELATATATDTDMWGNLPGDQNSNMLPTVDRLTALGAVGWMQON 417
 QY 610 RDVYLOGPIMAKIPIHTDGHFSPPLMGFGKLNKPPQILKNTPVANPAESATKFEAS 669
 DB 418 RDIYQGPIMAKIPHTDGHFSPPLMGFGKLNKPPQILKNTPVANPAESATKFEAS 477
 QY 670 FITQYSTQVSEIEMELQKENSCKMNPVQYTSNYSKASAVDFTVNNGLYTEPRPIGT 729
 DB 478 FITQYSTQVSEIEMELQKENSCKMNPVQYTSNYSKASAVDFTVNNGLYTEPRPIGT 537
 QY 730 RYLTRPL 736
 DB 538 RYLTHHL 544

RESULT 4
 US-08-856-841-22
 ; Sequence 22, Application US/08856841
 ; Patent No. 6274307
 ; GENERAL INFORMATION:
 ; APPLICANT: ERWIN SOUTSCHEK
 ; APPLICANT: MANFRED WOTZ
 ; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
 ; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
 ; STREET: 99 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10016
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" FLOPPY DISC
 ; COMPUTER: AT&T - IBM COMPATIBLE
 ; OPERATING SYSTEM: MS-DOS Version 6.2
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/856,841
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/214,658
 ; FILING DATE: 16-MARCH-1994
 ; APPLICATION NUMBER: US 07/917,096
 ; FILING DATE: 4-AUGUST-1992
 ; APPLICATION NUMBER: PCT/DE91/00106
 ; FILING DATE: 8-FEBRUARY-1991
 ; APPLICATION NUMBER: DE40038262
 ; FILING DATE: 8-FEBRUARY-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROBINSON, WILLIAM R.
 ; REGISTRATION NUMBER: 27,224
 ; REFERENCE/DOCKET NUMBER: LKR-9222-A
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-3355
 TELEFAX: (212) 557-5635
 TELE: NONE
 INFORMATION FOR SEQ ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 543
 TYPE: AMINO ACID
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PEPTIDE
 DESCRIPTION: N/A
 HYPOTHEICAL: N/A
 ANTI-SENSE: N/A
 FRAGMENT TYPE: INTERNAL
 ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
 ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONISUM)
 IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
 POSITION IN GENOME: N/A
 FEATURE:
 NAME/KEY: N/A
 LOCATION: N/A
 IDENTIFICATION METHOD: amino acid analysis and
 IDENTIFICATION METHOD: mass spectrometry
 OTHER INFORMATION:
 PUBLICATION INFORMATION:
 AUTHORS: COSSART, Y.B.
 AUTHORS: FIELD, A.M.
 AUTHORS: CANT, B.
 AUTHORS: WIDOWS, D.
 TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
 JOURNAL: LANCET
 VOLUME: 1
 ISSUE:
 PAGES: 72 - 73
 DATE: 1975
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO: 22:
 US-08-856-841-22

Query Match 12.0%; Score 479.5; DB 4; Length 543;
 Best Local Similarity 27.1%; Pred. No. 1,7e-33;
 Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16;

QY 212 ADNNEGADGVGANS--GNHCDSTWLGDRYITSTRTMALPTNNMLYKOISSASGASN 269
 DB 16 AEAATGAGGGGNSVSMSEGAATFSANVTCTFSRQFLIPYDREHYKVFSPASSCHN 75
 QY 270 D-----NHFGYSTPMGVFEDFNRFCHSPSPDMOQLINNMGFPRKRLNFTLFTN 319
 DB 76 ASGKEAKVCTISPIMGYSTPMKRYLDNALNLFSPDFQHLLENYSIADALTITISEI 135
 QY 320 QVKEVT--INDGVTTIANLNTSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMIPKGY 377
 DB 136 AVQDVTDKTGGGV-QYTDSTTGLCMVLDHEVYKPYVLGGQDTLAPBLPIWYVPPQVA 194
 QY 378 YLTIAN-GSQAVG-----RSSFYCLEYPPSQMLRTGNNFTSYTPEEVPFSSYAH 427
 DB 195 YLTVGQVNTQGISGDSKTLASESAFYVLEHSSFQILGTGTASMSYKPPVPEMLBGC 254
 QY 428 SOSLDLNPRLDQYLYLNRQNGSGAONKDLFSRGSFAGMSQVQPKMLPGPCYRQ 487
 DB 255 SQHFYEMNPL--YSGRLGVPDTLGDPRFSL---THEDHAIQQVFMFGPLVNSV 306
 QY 488 RVSKTIDNNNSNFTWTGASKYNLNGRESIINPG-TAMASHKDEDEKFFPMGSMIFGKE 546
 DB 307 STEGDSNVTGAKALTLGSLTGSQNTRLSLRGVPSQPHNMDTQXYTGIAISHGOT 366
 QY 547 SAGASNTALDNV-----MTDEEIKATNPVATERFGVAVNFQSSSTDPTAGDVHANG 599
 DB 367 TYG---NAEDKEYQGVGRFPNEKEQLKQLOGLNMHTY-----FNKGQOYTDQIE-R 416
 QY 600 GALPGWVQDRDVLVGLPIWAKIPIHTDGHFHPB-PLMGFGKLNKPPQILKNTPVANPA 658

Db 417 PLMTGWNRAHAYESQWLSKIPNLDSFKTQFALGQWGLHQPPI 465
Qy 659 PAEFAITFASFTIOYSTGVSELEML-OKENSKRNPE 658
Db 466 -----FLKQYAVGIMVTMTFTLGPBKATGRWNPQ 495

RESULT 5
Sequence 16, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LXR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 500
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y. E.

AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE: 1
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES: 1982
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
JOURNAL: GLUTATHIONE S. TRANSFERASE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
US-08-856-841-16
Query Match 10.8%; Score 430.5; DB 4; Length 500;
Best Local Similarity 29.5%; Pred. No. 2.9e-29;
Matches 122; Conservative 45; Mismatches 170; Indels 77; Gaps 10;
Qy 47 LPQYKYLGPFGNGLDKGEPVVAADDAALHEDKAYDQQLKAGDNFYLRNHNADAEFORLOE 106
Db 133 LPQTNVYVGPENELQAGPQSAVDSARIDFRYSQAKLGINRYTHMTVADELLKNIKN 192
Qy 107 DTSPFGNLRGAVQAKRVLEPLGLVEGAKTAPGKKRPVEGSPQEPDSSSGTGTGQOP 166
Db 193 ETGFOAQQVVDYF-----TLKGAAPVAPHPQ-----GSLPEYV 225
Qy 167 AKKRLNPGQGTGSESVDDPPLGEPATPAVAGPTTMAAGGAPMADNNEGADGVGNASG 226
Db 226 AYNA-----SEKPSMTSVNSAASSTGA-----GGG-----GNSVYS 258
Qy 227 NMHCDSTWLDGRVITSTRTWALPTYNNHLYKQISSASTGASND-----NHYFGY 276
Db 259 MMEBGAFTFANSVTCFTRQFLIPYDPEHHYKVFSPAASCHNAGSKAKVCTISPIMGY 318
Qy 277 STPWGTFDTRFRFCHSPRDMQRLINNNGFRKRLNFKLFNIQVEVT--TNDGYTTTA 334
Db 319 STEPWYLDFAVALNLFSPSPLEFQHLINYSIAADALTVTISETAVADVDTKGGGV-QVT 377
Qy 335 NNTSTVQVPSDEYQLPVYLGAHQGCLPPFPADVFMIPQGYLTLNN-GSQAVG----- 389
Db 378 DSTTGRLCMLVDHEYKPYVYLGQGGDTLAPBLPIWYFPQIAYLVGVNNTIGIGDSK 437
Qy 390 -----RSSFYCLEYFPSOMLRGNNFTFSYTFEEVDFHSSVYASQSLDRLMPL 438

Db 438 KLASSESAFYVLEHSSFOLLGTGTASMSYKFPVPPEPNELEGCSQHFEYEMYNPL 491

RESULT 6
US-08-856-841-18
Sequence 18, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-8635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA

JOURNAL: LANCET
VOLUME: I
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 18:
US-08-856-841-18

Query Match 10.8%; Score 430.5; DB 4; Length 501;
Best Local Similarity 29.5%; Pred. No. 2,9e-29;
Matches 122; Conservative 45; Mismatches 170; Indels 77; Gaps 10;

QY 47 LPQKTYLGPNGLDKGEVPAADAALEHDKAYDQKAGDNYLYRNHADAEPQERLQE 106
DB 134 LPQKTYVGPNGELQAGPQSAVDSAAIRIHFYSQAKLGINPYTWVADSELKNIK 193
QY 107 DTSEGGNLGRAVQAKRVLEPLGLVEBGAKTAPGKKRPVQSPQEPDSSGIGKTGQCP 166
DB 194 ETGFQAVVQVYF-----TLKGAAPVAHQ-----GSLPEVP 226
QY 167 AKRLNFGQDSESVDPQPLGEPATPAVGPPTMASGGA PMADNNEGADGVNASG 226
DB 227 AYNA-----SEKYPMSVNSAASATGA-----GGS-----GNSVKS 259
QY 227 NHCDSTWLDGRVITTTSTRTWALPTNNHLYKOISSASTGASND-----NHYPGY 276
DB 260 MMSGATFASANSYCTFSRQFLIPYDPEHHYKVFSPASSCHNAGSKAKVCTISPMGY 319
QY 277 STPGVYFENRPHCHSPRDMORLNNNGFRPKRLNFKLNFVQVEVT--TNDGYTTA 334
DB 320 STPWYLDENALNLFSPLEPQHLIENYGIADALVTTSIAVADVDTKGTGV-QVT 378
QY 335 NNLSTVQVPSDSEYQLPVYLGSAGHQCCLPPPADVFMTPOXYGLTLNN-GSQAVG---- 389
DB 379 DSTTGLCLMVDHKKYKPYVLGGGQDTLAPELIWIYFPQVXLYLVGDVNTGIGSDSK 438
QY 390 -----RSSPFCLEYPSPQMLRTGNNTFTSYTEEPVPHSSVAHSGSLDRMLNPL 438
DB 439 KLASSESAFYVLEHSSFOLLGTGTASMSYKFPVPPEPNELEGCSQHFEYEMYNPL 492

RESULT 7
US-08-856-841-19
Sequence 19, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994

Query Match	10.8%	Score 429.5	DB 4	Length 486
Best Local Similarity	28.8%	Pred. No. 3.4e-29		
Matches	120	Conservative 45	Mismatches 169	Indels 83
			Gaps	7

Qy	47	LPGYKTLGPFNGLDKGEFVNADAALBHDKAYDQOLKAGDPYLRYNHDAEFGERLOE	106
Db	125	LPGTIVYVPGNIIQAGPPQSAVDASARIRIDPFYSOLAKGIPIYHTMTVADEELIKIN	184
Qy	107	DTSPCGNIGRAVFOAK---KRYLBEGLVEBEAKTAPGKKRPVEBSPQEDSSSIGTKTG	163
Db	185	ETGFAQVAVKKYFTLKGAAPRAHFQGSIPPEYPAVNASEKPSMTSVNSAEKSTAGGGG	244
Qy	164	QQPARKRLNFGQTGDSSESVPDQPLGEPAPATPAVAGPTTMA5GGAPADNNEGADYGN	223
Db	245	SNPVYSM-----	251
Qy	224	ASGNMHCDSTWLGDRVITTTSTRTMALPTYNNHLVYQISSASTGASND-----NHV	273
Db	252	-----WSEBATESANSVTCFSPHOFILPYDPEHHYVFPSPA5SCHNAGSEAKVCTISPI	307
Qy	274	FGYSTPMGVPPDNNRHCHFSPPMDQRLINNMMGPRPKLNFGLPIYQKEVY--INDGVY	331

RESULT 8
 US-08-856-841-20
 Sequence 20, Application US/08856841
 Patent No 6274307
 GENERAL INFORMATION:
 APPLICANT: ERWIN SOUTSCHER;
 APPLICANT: MANFRED MOTZ
 TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
 TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
 NUMBER OF SEQUENCES: 28
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
 STREET: 99 PARK AVENUE
 CITY: NEW YORK
 STATE: NY
 COUNTRY: USA
 ZIP: 10016
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" FLOPPY DISC
 COMPUTER: AT&T - IBM COMPATIBLE
 OPERATING SYSTEM: MS-DOS Version 6.2
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/856,841
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/214,658
 FILING DATE: 16-MARCH-1994
 APPLICATION NUMBER: US 07/917,096
 FILING DATE: 4-AUGUST-1992
 APPLICATION NUMBER: PCT/DE91/00106
 FILING DATE: 8-FEBRUARY-1991
 APPLICATION NUMBER: DE40038262
 FILING DATE: 8-FEBRUARY-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: ROBINSON, WILLIAM R.
 REGISTRATION NUMBER: 27,224
 REFERENCE/DOCKET NUMBER: LKN-9222-A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-3355
 TELEFAX: (212) 557-5635
 TELEX: NONE
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 415
 TYPE: AMINO ACID
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PEPTIDE
 DESCRIPTION: PEPTIDE
 HYPOTHEICAL: N/A
 ANTI-SENSE: N/A
 FRAGMENT TYPE: INTERNAL
 ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
 ORIGINAL SOURCE: INFECTION (REYTHEM INFECTIONUM)
 IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
 POSITION IN GENOME: N/A
 FEATURE:
 NAME/KEY: N/A
 LOCATION: N/A
 IDENTIFICATION METHOD: amino acid analysis and
 IDENTIFICATION METHOD: mass spectrometry

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS: COSART, Y.E.

AUTHORS: FIELD, A.M.

AUTHORS: CANT, B.

AUTHORS: WIDOWS, D.

TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA

JOURNAL: LANCET

VOLUME: 1

ISSUE: 1

PAGES: 72 - 73

DATE: 1975

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO: 20:

US-08-856-841-20

Query Match 8.2%; Score 326; DB 4; Length 415;

Best Local Similarity 26.0%; Pred. No. 3.2e-20;

Matches 101; Conservative 66; Mismatches 180; Indels 42; Gaps 13;

QY 330 VTTIANNLSTVOYFSDSEYQLPVLSAAGCLPPPADVFMIPQYGLTLNN-GSOAV 388

Db 1 MTWITPLHACMLV--DHEKYKPYVLGGQDITLAPLPIWVYPPQYAVLTVGDVNTQGI 58

QY 389 G-----RSSFYCLEYFPSCMLRTGNNFTSTYEEVPHSSYAHSGSLRLNPLI 439

Db 59 SGDSKSLASESAFYVLEHSSFOLLGTGTASMSYKPPPPPELLECSOHFEMVPL- 117

QY 440 DOYLVIYINRTQNGSAGNDLLEFRSGPAGMSVQPGNWLPGPCYRQGVSKTKTDNNNS 459

Db 118 --YGSRLGVDTLGGDPKFRSL-----THEDHAIQPNFPGPLVNSVSTKEDGSNTGA 170

QY 500 NFTWTGASKYNLNGRESIINPG-TAMASHKODEDKFPMSGVMIFGESAGASNTALDNV 558

Db 171 GKALTGTSTGSONTRISLRPGVSPYHHMDTKYTGNAISHGQTYG--NAEDKE 227

QY 559 -----MTDEBKATNPVATERFGTVAVNFSSSTDPATGVDHAMGLPCKVWQDRD 611

Db 228 YQGVGVGRFPEKQKQLOGLNMHTY-----FENKGTQOYVQIE-RPLMVSQVWNRRA 280

QY 612 VYLGPIWAKIPHTDGHFHS-PLMGFGKLNPPPOLINNTPPANPAFSAATKPSAF 670

Db 281 LHYESQIWSKIPINDDSFKTQFALGSMGLHOPPPQFLK--ILPESGPIGKISMGITT 338

QY 671 ITQYSTQGVSRLEWEL-OKENSKRANPE 698

Db 339 LVQYAVGIMVTMTFKLGPRAKATGRANPQ 367

RESULT 9

US-08-856-841-13

Sequence 13, Application US/08856841

Patent No. 6274307

GENERAL INFORMATION:

APPLICANT: ERWIN SOUTSCHEK

APPLICANT: MANFRED MOTZ

TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES

TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY

STREET: 99 PARK AVENUE

CITY: NEW YORK

STATE: NY

COUNTRY: USA

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" FLOPPY DISC

COMPUTER: AT&T - IBM COMPATIBLE

OPERATING SYSTEM: MS-DOS Version 6.2

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/856,841

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/214,658

FILING DATE: 16-MARCH-1994

APPLICATION NUMBER: US 07/917,096

FILING DATE: 4-AUGUST-1992

APPLICATION NUMBER: PCT/DE91/00106

FILING DATE: 8-FEBRUARY-1991

APPLICATION NUMBER: DE40038262

FILING DATE: 8-FEBRUARY-1990

ATTORNEY/AGENT INFORMATION:

NAME: ROBINSON, WILLIAM R.

REGISTRATION NUMBER: 27,224

REFERENCE/DOCKET NUMBER: LKR-9222-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-3355

TELEFAX: (212) 557-5635

TELEX: NONE

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 395

TYPE: AMINO ACID

TOPOLOGY: LINEAR

MOLECULE TYPE: PEPTIDE

DESCRIPTION: N/A

HYPOTHETICAL: N/A

ANTI-SENSE: N/A

FRAGMENT TYPE: INTERNAL

ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE

ORIGINAL SOURCE: INFECTION (ERTHERIA INFECTION)

IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE

POSITION IN GENOME: N/A

FEATURE:

NAME/KEY: N/A

LOCATION: N/A

IDENTIFICATION METHOD: amino acid analysis and

IDENTIFICATION METHOD: mass spectrometry

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS: COSART, Y.E.

AUTHORS: CANT, B.

AUTHORS: WIDOWS, D.

TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA

JOURNAL: LANCET

VOLUME: 1

ISSUE: 1

PAGES: 72 - 73

DATE: 1975

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO: 13:

PUBLICATION INFORMATION:

AUTHORS: SMITH, D.B.

AUTHORS: JOHNSON, K.S.

TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME: 67
ISSUE: 1
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 13:
US-08-856-841-13

Query Match 8.0%; Score 318; DB 4; Length 395;
Best Local Similarity 25.4%; Pred. No. 1.5e-19;
Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;

QY 330 VTTIANNLTSTVQVFSDEYOLPYVLSAGHQCCLPPFADVFMIPQYGLTLANN-GSQAV 388
DB 1 MTMTTPSLHACMLV--DHEYKPYVLGGQODTLAPELPIWYFPQYAVLTVGQVNTQGI 58
QY 389 G-----RSSFYCLEYFPSPQMLRTGNFTSYTEEEVPHSSVAHSQSLDLRLANPLI 439
DB 59 SGDSKTLASBSAFAVLEHSSFFOLLGTGTASMSYKFPFPPENILGSCSHFEMYNPL- 117
QY 440 DQVLYINRTNONGSAGNKKLLRSRSPAGMSVQPKMLRPGCYROORVSKTKTDNNNS 499
DB 118 --YQSRIGVPTLTGDDPKFRSL---THEDHAIQPOFMFKPLVNSVSTYKGGSSVTGA 170
QY 500 NFWTGASKYVINGRESIINPG--TAMASHKDEDEKFFPMGSMVLFKESAGASNTALDNV 558
DB 171 GKALTLGSLTGSQNTRLSLRGPVSQPHMDTKYVIGVIAIHGQTTG---NADKE 227
QY 559 -----MTDEEIKATNPVATERFGTVAVNFSSSTDPATYGVNHAMGALPGVWQDRD 611
DB 228 YQGVGFRPEKEQKLOGLMHTY-----FPRKGTQYTDIIR-RPLAVGSVMNRRA 280
QY 612 VYLGQPIWAKIPHTDGHFHPG--PLMGGRGLKNPPQILIKNTVPVAPNPAEFSATKFSF 670
DB 281 LHYESQMSKLPNLDSEFKTPAALGGWGLHQPDDI-----F 318
QY 671 ITQYSTGVSVIEIEMEL-QKENSXRMPE 698
DB 319 LKQYAVGIMTYTMTFKLGPRAKATGRMNQ 347

RESULT 10
US-08-856-841-14
Sequence 14, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 264
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.B.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE: 72 - 73
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE

VOLUME: 67
ISSUE: 31 - 40
PAGES: 1988
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
US-08-856-841-14

Query Match 7.8%; Score 310; DB 4; Length 264;
Best Local Similarity 31.3%; Pred. No. 3.9e-19;
Matches 78; Conservative 33; Mismatches 104; Indels 34; Gaps 6;

QY 194 TPAVAGPTTASGGAGFAMNNEGADGVNAGSNMHCDSWMLGRVYTTSTRMALPTYN 253
DB 5 TELAAEASGTAGGG-----GSNSVKSNMSEAGATFASNSVCTFSRQPLPYDP 53
QY 254 NMLYKOISSASTGASND-----NHYFGYSTPMGYDFNRFCHFSRDMORLNN 303
DB 54 EHHYKVFSPASSCHNAGSKAKVCTISPLMGYSTPRYIDFALNLFPSLEFQHLIEN 113
QY 304 NMGFRPRLNFKLPNIQVKEVT--TNDGVTTIANNLTSVQVFSDESYQLPYLGSAGHG 361
DB 114 YGSIAPDALVTITSEIAVKDVTDKTGGV-QVTDSTTGRCLMVDHKKYPYLGOGODT 172
QY 362 CLPEFPADVFMIOGYLTINN-GSQAVG-----RSFYCLEFSPOMLRTGNFT 411
DB 173 LAELPLWVFPPOYALTVGDVNTQGISDSKSLASESAFYVLEHSSFOLLGTGTAS 232
QY 412 FSYTFEVP 420
DB 233 MSYKFPVP 241

RESULT 11

US-08-856-841-21
Sequence 21, Application US/08856841
Patent No. 6274307

GENERAL INFORMATION:

APPLICANT: ERWIN SCUTSCHER
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SOURCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224

REFERENCE/DOCKET NUMBER: LKR-9222-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-3355

TELEFAX: (212) 557-5635

TELEX: NONE

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 398

TYPE: AMINO ACID

TOPOLOGY: LINEAR

MOLECULE TYPE:

DESCRIPTION: PEPTIDE

HYPOTHETICAL: N/A

ANTI-SENSE: N/A

FRAGMENT TYPE: INTERNAL

ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE

IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE

POSITION IN GENOME: N/A

FEATURE:

NAME/KEY: N/A

LOCATION: N/A

IDENTIFICATION METHOD: amino acid analysis and

IDENTIFICATION METHOD: mass spectrometry

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS: COSSART, Y.E.

AUTHORS: FIELD, A.M.

AUTHORS: CANT, B.

AUTHORS: WIDOWS, D.

TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA

JOURNAL: LANCET

VOLUME: I

ISSUE: I

PAGES: 72 - 73

DATE: 1975

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO: 21:

US-08-856-841-21

Query Match 7.6%; Score 303; DB 4; Length 398;

Best Local Similarity 25.4%; Pred. No. 3.1e-18;

Matches 96; Conservative 48; Mismatches 168; Indels 66; Gaps 12;

QY 346 DSEYQLPYLGSAGHGCCLPEFPADVFMIOGYLTINN-GSQAVG-----RSFYC 395
DB 4 DHEKYPYVGGOGDTLAPLPIWVFPPOYALTVGDVNTQGISDSKSLASESAFYV 63
QY 396 LEFPSPOMLRTGNFTFSYTFEVPFPHSSYAHQSOLDRMLNPLIDYLYIARTQNGS 455
DB 64 LEHSSFOLLGTGTATWSTKFPVPPEINEGCSQHFEYEMNPL--YGRIGVPTLGGD 120
QY 456 AQNKDLFGRSPAGMSVQKWLPGPCYRQORVSTKTDNNNSNFTWGASKYINLNGRE 515
DB 121 PKRSL-----THEDALIQPNFMFPLVNSVSTKGDSNAGALGLSTGTQNR 175
QY 516 SIINRG-TMASHKDEDEKFPFMSGVMIFGKSAGASNTALDNV-----MITDEEIK 567
DB 176 ISLRPGVPOPYHMDTDKYVTGINAISHGQTYG--NAEDKEYQGQVGRFPNEKEQLK 232
QY 568 ATPVATERFGTVAVNFOSSSTDPAFGVHAMALPGMWQORDVYLOGPIMAKIPHTG 627
DB 233 QLOGLMHTY-----FKNKGTQYTDQLE-RPLWGSVNNRRALHYESQLWSKITNLD 285
QY 628 HFFPS-PLMGFGKLPKPPQILIKNTVPANPAPBSATKFAFITQYSTGVSEIENE 686
DB 286 SFKTPALAGWGLHQPPIQ-----FKYHVKVGOLEVLNQWE 323
QY 687 LQKNSK-----RWNP 697
DB 324 LLEFNNPWELOLHNGP 341

RESULT 12
US-08-856-841-17
Sequence 17, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHER
APPLICANT: MANFRED MORTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 387
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIONIS)
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE:

PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 17:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 15:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 17:
US-08-856-841-17

Query Match 7.1%; Score 284; DB 4; Length 387;
Best Local Similarity 28.0%; Pred. No. 1,4e-16;
Matches 82; Conservative 28; Mismatches 119; Indels 64; Gaps 6;

QY	47	LPGLKYLGPVNGLDKGEVPAADAALEHDKAYDQQLKAGDNYLYKNHADAEPQERLQ	106
DB	133	LPGLTYVGPVNGELQAGPQGAVDAAIRIHDFRSQAKGGINVTYHTWTVADEILKNIXN	192
QY	107	DTSEGGNLGRAVFOAKRVLEPLGLVEBAKTAPEGKRPVEQSPQEPDSSGIGTKGOOP	166
DB	193	ETGFOAQQVVKDYF-----TLKGAAPVAHFQ-----GSLPEVP	225
QY	167	AKRLNFGQTGDSSEVDPDQPLGEPATPAVAGPTTMAAGGAPMADNNEGADGVGNASG	226
DB	226	AYNA-----SEKYPSSMTSVNSAEASTGA-----GGSNSVKS	258
QY	227	NMHCDSTWLDGRVITTSRTWALPTNNHLYKQISSASTGASND-----NHIFGY	276
DB	259	MMEGATFASANSYTCFTRQFLIPYDEHHYKVFSPASSCHNAGSEAKVCTISPIMGY	318
QY	277	STPMGVEDENRFHCHSPRDMORLNNMGFRPKRLNFKLNIQVKEVTINDG	329
DB	319	STPMRYLDENALNDPSPLEFQHLIENYGSIALDALTTVTSIAVADVTIKTG	371

RESULT 13
5223424-13
Patent No. 5223424
APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
RICHARD D.
TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
ACID SEQUENCE
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/225, 032

FILING DATE: 27-JUL-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 78, 519
 FILING DATE: 27-JUL-1987
 APPLICATION NUMBER: 933,107
 FILING DATE: 20-NOV-1986
 APPLICATION NUMBER: 902,887
 FILING DATE: 02-SEP-1986
 APPLICATION NUMBER: 887,140
 FILING DATE: 17-JUL-1986
 APPLICATION NUMBER: 823,102
 FILING DATE: 27-JAN-1986
 APPLICATION NUMBER: 773,430
 FILING DATE: 06-SEP-1985
 SEQ ID NO:13:
 LENGTH: 579
 5223424-13

Query Match 5.4%; Score 216.5; DB 6; Length 579;
 Best Local Similarity 22.6%; Pred. No. 2,2e-10;
 Matches 143; Conservative 84; Mismatches 250; Indels 157; Gaps 33;

179 SESVDPDPLGEPATPAVPTTMAAGGAPMADNNEGADVGNASGNMH--CDSTWIG 236
 2 SEVEQHNPINAGTEL--SATGNESGGGGGGG--GARGAGVGVSTGSENNQTEFOYL 56
 237 D--RYITSTRFALPTVNNHLYKQIS--SASTGAS--NDNHFGYSTPGYFDFR 287
 57 EGVLRITTAHSAIRLHINMPHEHYKRIHVLNSESVAQGVQVADATQVTPSLIDRA 116
 288 FHCHFSRDMQRLINNNWGFPRKLNFKLNIQVEVT--TNDGVTTIANLITSTVOVF 344
 117 WGVWFPAWMQILSNMTEINLVSFEQEIFNVVLKTTTSATSPPSKYINNDLTASLMVA 176
 345 SDSEYOLPYVLSAHOGLCPPRPADVPMIPQYGL-----TLNNGSQAVGRS--- 391
 177 LDTNNLPTPAAPRSETLGFYPMLPKPYQRYVLSCLRNLPPTVYTGSSQGITDSIQ 236
 392 -----SPFCLF-YFPSONLRTGNF-TFSYTFEEVPHSSVAHSQSLRLNPLLDQYI 443
 237 GLHSDIMEFTIENAVVHILRTGDEFSIGIYHDTYPL--KLTHSWQTNK----- 284
 444 YVILNRTONOSGSAQNKLDFSRSPPAGMSVQP-----KNMLPGCYRQ--QRYSKTK 493
 285 -----SLGLPRKVLTEPTTBEDQHGTLPKANTKRGHYQTINN 324
 494 TD-----NNNSFTWTGASKYINLNGRESIINP--GTAMASHKDEDEKFFPMG 539
 325 TEATLAPQAVGYNTPEYMFEDYSNGSPF-----LTPIVPADTQYVDE---PNGA 372
 540 VNI FGKESAGASTALDNVITDEBEIKATNP-----VATERFGIVA--VNPQSSS--- 588
 373 IRETMGYOGLHTTS-----SOELERYTFNPOSCKGRAPKQOPNOAPLNTLNTNGTL 426
 589 --TDPATG--DVHAMGALP-----GMWQDRDYLLOGPWAKIPHTD--GHFSPSLMG 637
 427 LPSDPIGKSNKHFMTLNTYGLTALNNTAPVFPNGQIMDKLDDDKRLH---VTPA 483
 638 FGLK-NPPQILIKNTVPANPAEFA--TKFASPTTQYSTGVSEVEIEMELOKENSKR 694
 484 FVCKKNPQQLFYKIP--NLTFDENADSPOQPRIITD--SNFMKGTLLFTAKMRSSNN 539
 695 WNPVEVYTSNYAKSANVDFVNDNGLYTEBRPIG 728
 540 WNPFOQHT-----TAENIRKYL-PTNIG 562

RESULT 14
 US-09-022-949-2
 ; Sequence 2, Application US/09022949
 ; Patent No. 6187759
 ; GENERAL INFORMATION:
 ; APPLICANT: Tarpey, Ian

APPLICANT: Greenwood, Neil
 TITLE OF INVENTION: Canine parvovirus DNA vaccination
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Akzo No. 6187759e1 Patent Dept.
 STREET: 1300 Piccard Drive, Suite 206
 CITY: Rockville
 STATE: Maryland
 COUNTRY: US
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30 (ERO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/022,949
 FILING DATE: 11-FEB-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gornley, Mary E.
 REGISTRATION NUMBER: 34,409
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-948-7400
 TELEFAX: 301-948-9751
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 584 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-022-949-2

Query Match 4.8%; Score 192; DB 4; Length 584;
 Best Local Similarity 19.9%; Pred. No. 3.1e-08;
 Matches 128; Conservative 100; Mismatches 232; Indels 184; Gaps 32;

186 OPLGERPA--TAAVPTTMAAGGAPMADNNEGADVGNASGNMH-----CDSTWIG 236
 7 QDPGQPAVRNBRATSGSGGGG-----GGSGGVISITGTENNQTEFKLENGW- 59
 237 DRVITSTRFALPTVNNHLYKQI-----SSASTGASNDNHVFGYSTPGYFDFRPH 289
 60 -ETIANSRLVHLMNPESNRYRVVNNLDKTAVGNMALLDTHAQIVTPMSLVDANANG 118
 119 VWFNPDQGLIVNMSHLVSEFQEIFNVVLKTVSESATOPTKYVNNNDLTASLMVALD 178
 290 CHFSRDMQRLINNNWGFPRKLNFKLNIQVEVT--TNDGVTTIANLITSTVOVFS 346
 119 VWFNPDQGLIVNMSHLVSEFQEIFNVVLKTVSESATOPTKYVNNNDLTASLMVALD 178
 347 SEYOLPYVLSAHOGL-----PPPA-----DVMIQYGLITLNGSQAVGRS- 391
 179 SNNMTEFTPAARSETLGFYPMKPTIPFWRYYFOWDRILPSH-----TGTSGPTVI 232
 392 -----SPFCLF-YFPSONLRTGNF-TFSYTFEEVPHSSVAHSQSLRLNPLLD 440
 233 YGTDPDVQVFTIENSVAHLRTGDEPATGTFFDCKP--CRLTHQTNRALG--LP 288
 441 QVLYILNRTONOS--GSAQNK-----DLFSRGSPPAGMSVQPKNMLPG 481
 289 PFLNLSQSEGGTNFGYIGVQDKRGVYTGOMNTVYTBATIMRPAEVGS-----A 340
 482 PCYRQORVSK-----TTDDNNSFTWTGASKY--NUNGRESIINPCT--- 522
 341 PYSFASSTQGPFPKPIAAGGAGQTDENQAA--DGDPRYAFGRHQGKQTTTGTTPR 397
 523 -AMASHKDEDEKFFPMGVMIFGKESAGASTALD--NVMITDEBEIKATNPVATERFGTV 580
 398 FTYIAHQD-----GRYPEGWIONINPFLVTDNVLLPTDPIG-----GRT 440
 581 AVNFQSSSTDPAIGDVHAMGALPGWVQDRDYLLOGPWAKIPHTDGHFSPSLMGFGL 640
 441 GINY--TNIFNTYGLTALNVP-----PYVNGQIMDKKFPD-----L 478

QY 641 K-----NPEPOLIKTPVPAN---PPEFSATKTFASPTIOYSGQVSVEIE 684
DB 479 KPELHVNPAPVCCNCCGQLFVKAPNLTNEYDPDASANSR-----LVTVSDFMWKGKLV 534
QY 685 WELOKENSKEKNEPEVQYTSNVAKSANVDFTVNNGLYTEPRPIG 728
DB 535 FKAKLRASHMTWNPICQMSIN-----IDNQFNV-PSNIG 567

RESULT 15
US-08-856-841-9
; Sequence 9, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: ERWIN SOUTSCHER
; APPLICANT: MANFRED MOTZ
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
; STREET: 99 PARK AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISC
; COMPUTER: AT&T - IBM COMPATIBLE
; OPERATING SYSTEM: MS-DOS Version 6.2
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856,841
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,658
; FILING DATE: 16-MARCH-1994
; APPLICATION NUMBER: US 07/917,096
; FILING DATE: 4-AUGUST-1992
; APPLICATION NUMBER: PCT/DE91/00106
; FILING DATE: 8-FEBRUARY-1991
; APPLICATION NUMBER: DE40038262
; FILING DATE: 8-FEBRUARY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINSON, WILLIAM R.
; REGISTRATION NUMBER: 27,224
; REFERENCE/DOCKET NUMBER: LKR-9222-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-3355
; TELEFAX: (212) 557-5635
; TELEX: NONE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PEPTIDE
; DESCRIPTION: N/A
; HYPOTHETICAL: N/A
; ANTI-SENSE: N/A
; FRAGMENT TYPE: INTERNAL
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
; POSITION IN GENOME: N/A
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD: amino acid analysis and
; IDENTIFICATION METHOD: mass spectrometry
; OTHER INFORMATION:
; PUBLICATION INFORMATION:

QY 47 LPGYKYLGPENGIDKGPVNAADAALEHDKAYDQQLKAGDNFYLRNHNADAEFOERLOE 106
DB 126 LFGNIVVGPENELQAGPPOSADVSAARIDFRYSQAKLGINFTYHTVADBELLNKIKV 185
QY 107 DTSPGAGLRAVFOAK 122
DB 186 ETGFOAQVAVDYPTLK 201

Query Match 3.6%; Score 142; DB 4; Length 210;
Best Local Similarity 39.5%; Pred. No. 0.00015;
Matches 30; Conservative 10; Mismatches 36; Indels 0; Gaps 0;

US-08-856-841-9
; AUTHORS: COSSART, Y.E.
; AUTHORS: FIELD, A.M.
; AUTHORS: CANT, B.
; AUTHORS: WIDDOWS, D.
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
; JOURNAL: LANCET
; VOLUME: 1
; ISSUE:
; PAGES: 72 - 73
; DATE: 1975
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 9:
; PUBLICATION INFORMATION:
; AUTHORS: MANIATIS, T.
; AUTHORS: FRITSCH, E.F.
; AUTHORS: SAMBROOK, J.
; TITLE: MOLECULAR CLONING
; JOURNAL: COLD SPRING HARBOR, NY
; VOLUME:
; ISSUE:
; PAGES:
; DATE: 1982
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 9:
; PUBLICATION INFORMATION:
; AUTHORS: SMITH, D.B.
; AUTHORS: JOHNSON, K.S.
; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
; TITLE: GLUTATHIONE S. TRANSFERASE
; JOURNAL: GENE
; VOLUME:
; ISSUE: 67
; PAGES: 31 - 40
; DATE: 1988
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 9:

Search completed: July 17, 2003, 18:36:36
Job time: 20.1145 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 35.444 Seconds
(without alignments)
2466.245 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989

Sequence: 1 MAADGYLPDWLEBDNLSEGR.....NNGLYTEPPRIGTRVLRPL 736

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PC7US_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3402.5	85.3	735	US-10-038-972A-13	Sequence 13, Appl
2	3402.5	85.3	735	US-10-293-478-1	Sequence 1, Appl
3	2759.5	69.2	598	US-10-038-972A-14	Sequence 14, Appl
4	2481.5	62.2	533	US-10-038-972A-15	Sequence 15, Appl
5	1410	35.3	756	US-10-205-942-4	Sequence 4, Appl
6	119.5	3.0	1394	US-09-839-996-2	Sequence 2, Appl
7	119.5	3.0	1394	US-10-080-505-2	Sequence 7, Appl
8	118	3.0	2344	US-09-815-242-12713	Sequence 12713, A
9	117.5	2.9	1395	US-10-080-505-7	Sequence 7, Appl
10	117	2.9	1016	US-09-815-242-5845	Sequence 5845, Ap
11	116.5	2.9	5795	US-09-815-242-12610	Sequence 12610, A
12	115.5	2.9	410	US-10-022-809-2	Sequence 2, Appl
13	112	2.8	1441	US-10-170-682-3	Sequence 3, Appl
14	111.5	2.8	1265	US-09-862-027-80	Sequence 80, Appl
15	111	2.8	433	US-09-859-888-4	Sequence 4, Appl
16	111	2.8	2364	US-10-156-761-7834	Sequence 7834, Ap

17	110	2.8	938	US-10-029-217A-28	Sequence 28, Appl
18	110	2.8	938	US-10-029-217A-29	Sequence 29, Appl
19	109.5	2.7	1848	US-09-839-996-6	Sequence 6, Appl
20	109.5	2.7	1848	US-10-080-505-6	Sequence 6, Appl
21	109	2.7	1545	US-09-839-996-4	Sequence 4, Appl
22	109	2.7	1545	US-10-080-505-4	Sequence 4, Appl
23	108	2.7	1463	US-09-971-536-69	Sequence 69, Appl
24	106.5	2.7	331	US-09-996-561-1	Sequence 1, Appl
25	106.5	2.7	331	US-09-884-948-1	Sequence 1, Appl
26	106.5	2.7	331	US-10-112-488-5	Sequence 5, Appl
27	106.5	2.7	332	US-09-892-864A-2	Sequence 2, Appl
28	106.5	2.7	407	US-10-124-429-4	Sequence 4, Appl
29	106.5	2.7	407	US-10-112-488-13	Sequence 13, Appl
30	106.5	2.7	1023	US-10-137-866-200	Sequence 200, App
31	106.5	2.7	1023	US-10-146-726-200	Sequence 200, App
32	106.5	2.7	1023	US-10-146-727-200	Sequence 200, App
33	106.5	2.7	1023	US-10-146-788-200	Sequence 200, App
34	106.5	2.7	1023	US-10-152-380-200	Sequence 200, App
35	106.5	2.7	1023	US-10-153-934-200	Sequence 200, App
36	106.5	2.7	1023	US-10-028-072-200	Sequence 200, App
37	106.5	2.7	1023	US-10-121-049-200	Sequence 200, App
38	106.5	2.7	1023	US-10-123-904-200	Sequence 200, App
39	106.5	2.7	1023	US-10-140-470-200	Sequence 200, App
40	106.5	2.7	1023	US-10-175-746-200	Sequence 200, App
41	106.5	2.7	1023	US-10-176-918-200	Sequence 200, App
42	106.5	2.7	1023	US-10-176-921-200	Sequence 200, App
43	106.5	2.7	1023	US-10-137-865-200	Sequence 200, App
44	106.5	2.7	1023	US-10-140-474-200	Sequence 200, App
45	106.5	2.7	1023	US-10-142-431-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-10-038-972A-13
; Sequence 13, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 735
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VPI caspid protien
; US-10-038-972A-13

Query Match	85.3%	Score 3402.5;	DB 15;	Length 735;
Best Local Similarity	83.3%	Pred. No. 1.3e-281;		
Matches 613;	Conservative 51;	Mismatches 71;	Indels 1;	Gaps 1;
QY	1	MAADGYLPDWLEBDNLSEGRBWDLPKAPKPKRANQODDGRGLVLPGYKYLGPFGILD	60	
DB	1	MAADGYLPDWLEBDNLSEGRBWDLPKAPKPKRANQODDGRGLVLPGYKYLGPFGILD	60	
QY	61	KGPVNAADAAAEHDKAYDQOLKADNRYLRNHAADAEFOELQEDTSGGNGIGRAVFO	120	
DB	61	KGPVNAADAAAEHDKAYDQOLKADNRYLRNHAADAEFOELQEDTSGGNGIGRAVFO	120	
QY	121	AKKRVLEPLGLVEGAKTAPGKRPVPGSSGSGIGTKQOPAKKRLNFGOTCDSE	180	
DB	121	AKKRVLEPLGLVEGAKTAPGKRPVPGSSGSGIGTKQOPAKKRLNFGOTCDSE	180	
QY	181	SVDDPPLGPPAPSPGLGNTTATGSGAPMADNNEGALDVGNSGNGWCHDSTWMDRYI	240	
DB	181	SVDDPPLGPPAPSPGLGNTTATGSGAPMADNNEGALDVGNSGNGWCHDSTWMDRYI	240	

QY 241 TTSRTMALPTNNHLYKQISSASGASNDNHFGYSTPMGYPDPFNRFCHFSRDMQRL 300
DB 241 TTSRTMALPTNNHLYKQISSOS -GASNDNHFGYSTPMGYPDPFNRFCHFSRDMQRL 299
QY 301 INNMWGFPRKRLNFKLFINIQVEVTTNDGVTIANNLSTVOVPSDESEYOLPYVLGSAHQ 360
DB 300 INNMWGFPRKRLNFKLFINIQVEVTTNDGVTIANNLSTVOVPSDESEYOLPYVLGSAHQ 359
QY 361 GCLPPFPADVMTPOYGYLTLLNNGSOAVGRSSFYCLEFPPSOMLRTGNFTFSYTEEDVP 420
DB 360 GCLPPFPADVMTPOYGYLTLLNNGSOAVGRSSFYCLEFPPSOMLRTGNFTFSYTEEDVP 419
QY 421 FHSYASQSGLDRLMNPIDQYLYLNTQNGSQAQNKDLFSGSPAGMSVOPKXWLP 480
DB 420 FHSYASQSGLDRLMNPIDQYLYLNTQNGSQAQNKDLFSGSPAGMSVOPKXWLP 479
QY 481 GPCYRQORVSKTADNNNSFTWTGASKYLNGBRESIINPGTAMASHKDEDEKFFPMGCV 540
DB 480 GPCYRQORVSKTADNNNSFTWTGASKYLNGBRESIINPGTAMASHKDEDEKFFPMGCV 539
QY 541 MIFKESAGASNTALDNVMTDEEBIKATNPVATERFGTAVNFQSSSTDPAQGVHAG 600
DB 540 LIFKQSEKTNVDIEKMTDEEBIRTTNPVATEQVGSVSTLQGRNQAAATADVATQG 599
QY 601 ALPGWMODRDVYLOGPIMAKIPHTDGHFSPSLMGFGGLKNPPOILLIKNTVPANPST 660
DB 600 VLPQWMODRDVYLOGPIMAKIPHTDGHFSPSLMGFGGLKNPPOILLIKNTVPANPST 659
QY 661 EFSATKPFASFTIOYSTGVSEIEMELOKENSKRANPEVOYTSNYKASANDFTVNNGL 720
DB 660 TFSAKPFASFTIOYSTGVSEIEMELOKENSKRANPEIYQTSNYKASVANDFTVNNGV 719
QY 721 YTEPRPIGTRYLTRPL 736
DB 720 YSEBRPIGTRYLTRNL 735

RESULT 2
US-10-293-478-1

Sequence 1, Application US/10293478
Publication No. US20030078411A1
GENERAL INFORMATION:
APPLICANT: PATEL, SATI D.
TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
FILE REFERENCE: 38672
CURRENT APPLICATION NUMBER: US/10/293,478
CURRENT FILING DATE: 2002-11-14
PRIORITY APPLICATION NUMBER: US/09/321,589
PRIORITY FILING DATE: 1999-05-28
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1
LENGTH: 735
TYPE: PRT
ORGANISM: Adeno-associated virus
US-10-293-478-1

Query Match 85.3%; Score 3402.5; DB 15; Length 735;
Best Local Similarity 83.3%; Pred. No. 1.3e-281;
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps -1;

QY 1 MAAGGYLDMLTDLSEGISREWMDLKPGAPKPKANQKODDGRGLVPGYKYLGPFGLD 60
DB 1 MAAGGYLDMLTDLSEGISREWMDLKPGAPKPKAERKDDSRGLVPGYKYLGPFGLD 60
QY 61 KGEFVNADAAALBHDKAYDQOLKAGDNPYLRYNHADAEFOERLQEDTSGFSGNLGRAVQ 120
DB 61 KGEFVNADAAALBHDKAYDQOLKAGDNPYLRYNHADAEFOERLQEDTSGFSGNLGRAVQ 120
QY 121 AKKRVLEPLGLVEGAKTAPGKKRPVEOSPOEPDSSSIGTGTGQOPAKRLNFGQTDSE 180
DB 121 AKKRVLEPLGLVEGAKTAPGKKRPVEOSPOEPDSSSIGTGTGQOPAKRLNFGQTDSE 180

DB 121 AKKRVLEPLGLVEGAKTAPGKKRPVEOSPOEPDSSSIGTGTGQOPAKRLNFGQTDAD 180
QY 181 SVDPDQPLGEPATPAVGETTMASSGGAPMANNNGADGVNAGNWHCDSTWLDGRVI 240
DB 181 SVDPDQPLGEPATPAVGETTMASSGGAPMANNNGADGVNAGNWHCDSTWLDGRVI 240
QY 241 TTSRTMALPTNNHLYKQISSASGASNDNHFGYSTPMGYPDPFNRFCHFSRDMQRL 300
DB 241 TTSRTMALPTNNHLYKQISSOS -GASNDNHFGYSTPMGYPDPFNRFCHFSRDMQRL 299
QY 301 INNMWGFPRKRLNFKLFINIQVEVTTNDGVTIANNLSTVOVPSDESEYOLPYVLGSAHQ 360
DB 300 INNMWGFPRKRLNFKLFINIQVEVTTNDGVTIANNLSTVOVPSDESEYOLPYVLGSAHQ 359
QY 361 GCLPPFPADVMTPOYGYLTLLNNGSOAVGRSSFYCLEFPPSOMLRTGNFTFSYTEEDVP 420
DB 360 GCLPPFPADVMTPOYGYLTLLNNGSOAVGRSSFYCLEFPPSOMLRTGNFTFSYTEEDVP 419
QY 421 FHSYASQSGLDRLMNPIDQYLYLNTQNGSQAQNKDLFSGSPAGMSVOPKXWLP 480
DB 420 FHSYASQSGLDRLMNPIDQYLYLNTQNGSQAQNKDLFSGSPAGMSVOPKXWLP 479
QY 481 GPCYRQORVSKTADNNNSFTWTGASKYLNGBRESIINPGTAMASHKDEDEKFFPMGCV 540
DB 480 GPCYRQORVSKTADNNNSFTWTGASKYLNGBRESIINPGTAMASHKDEDEKFFPMGCV 539
QY 541 MIFKESAGASNTALDNVMTDEEBIKATNPVATERFGTAVNFQSSSTDPAQGVHAG 600
DB 540 LIFKQSEKTNVDIEKMTDEEBIRTTNPVATEQVGSVSTLQGRNQAAATADVATQG 599
QY 601 ALPGWMODRDVYLOGPIMAKIPHTDGHFSPSLMGFGGLKNPPOILLIKNTVPANPST 660
DB 600 VLPQWMODRDVYLOGPIMAKIPHTDGHFSPSLMGFGGLKNPPOILLIKNTVPANPST 659
QY 661 EFSATKPFASFTIOYSTGVSEIEMELOKENSKRANPEVOYTSNYKASANDFTVNNGL 720
DB 660 TFSAKPFASFTIOYSTGVSEIEMELOKENSKRANPEIYQTSNYKASVANDFTVNNGV 719
QY 721 YTEPRPIGTRYLTRPL 736
DB 720 YSEBRPIGTRYLTRNL 735

RESULT 3
US-10-038-972A-14

Sequence 14, Application US/10038972A
Publication No. US20020192823A1
GENERAL INFORMATION:
APPLICANT: J. Bartlett
TITLE OF INVENTION: AAV VECTORS AND METHODS
FILE REFERENCE: 28335/36966US
CURRENT APPLICATION NUMBER: US/10/038,972A
CURRENT FILING DATE: 2002-01-04
PRIORITY APPLICATION NUMBER: US 60/260,124
PRIORITY FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In version 3.1
SEQ ID NO 14
LENGTH: 598
TYPE: PRT
ORGANISM: adeno-associated virus 2 VP2 capsid protien
US-10-038-972A-14

Query Match 69.2%; Score 2759.5; DB 15; Length 598;
Best Local Similarity 82.8%; Pred. No. 7.5e-227;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APKKRPVEOSPOEPDSSSIGTGTGQOPAKRLNFGQTDSESVDPQPLGEPATPAV 198
DB 2 APKKRPVEOSPOEPDSSSIGTGTGQOPAKRLNFGQTDSESVDPQPLGEPATPAV 198
QY 199 GPTTMASSGGAPMANNNGADGVNAGNWHCDSTWLDGRVITTSRTMALPTNNHLYK 258
DB 199 GPTTMASSGGAPMANNNGADGVNAGNWHCDSTWLDGRVITTSRTMALPTNNHLYK 258

Db 62 GTTATATGAGPAMADNEGADVGNSSGNMHCSTWMDREVTITSTRTMAALPTNNHLYK 121
Qy 259 QISSASTGASNDNHGYGISTPAGYFPDNRPHCFSPRDORLLNNWNGFRPKLNFLEN 318
Db 122 QISSOS-GASNDNHGYGISTPAGYFPDNRPHCFSPRDORLLNNWNGFRPKLNFLEN 180
Qy 319 IQVKEVTTNDGVTIANNLSTVQVPSDSSEYOLPYVLGSAHQGLPPFPADVMIPOYGY 378
Db 181 IQVKEVTTNDGVTIANNLSTVQVPSDSSEYOLPYVLGSAHQGLPPFPADVMIPOYGY 240
Qy 379 LTLNNGSOAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVPHSSVAHSQSLDRMLNPL 438
Db 241 LTLNNGSOAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVPHSSVAHSQSLDRMLNPL 300
Qy 439 IDOYLTLARTQOSSGAQKDLIFSRGSPAGMSVOPKXMLBPCYRQOQVSTKTKDNN 488
Db 301 IDOYLTLARTQOSSGAQKDLIFSRGSPAGMSVOPKXMLBPCYRQOQVSTKTKDNN 360
Qy 499 SNTWTGASKYNNGRESINPGTAMASHKDEDEKFPFGSGWMI FGKESAGASNTALDNY 558
Db 361 SNTWTGASKYNNGRESINPGTAMASHKDEDEKFPFGSGWMI FGKESAGASNTALDNY 420
Qy 559 MITDEEIRKATNPVATERFCTVAVNFQSSSTDPATGVDHAMGALPGWMDRVDVYLGPI 618
Db 421 MITDEEIRKATNPVATERFCTVAVNFQSSSTDPATGVDHAMGALPGWMDRVDVYLGPI 480
Qy 619 WAKIPTHGHPSPMLGSGGLKNPPQILLIKTVPANPAEBSATKFSFTTQYSTGQ 678
Db 481 WAKIPTHGHPSPMLGSGGLKNPPQILLIKTVPANPAEBSATKFSFTTQYSTGQ 540
Qy 679 VSVIEIMELQKENSKRNNPEVOYTSNYSKASANDFTVDNNGLYTEPRPICTRYLTRPL 736
Db 541 VSVIEIMELQKENSKRNNPEVOYTSNYSKASANDFTVDNNGLYTEPRPICTRYLTRPL 598

RESULT 4

US-10-038-972A-15
Sequence 15, Application US/10038972A
Publication No. US20020192823A1
GENERAL INFORMATION:
APPLICANT: J. Bartlett
TITLE OF INVENTION: AAV VECTORS AND METHODS
FILE REFERENCE: 28335/3696US
CURRENT APPLICATION NUMBER: US/10/038, 972A
CURRENT FILING DATE: 2002-01-04
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 533
TYPE: PR
ORGANISM: adeno-associated virus 2 VP3 capsid protien
US-10-038-972A-15

Query Match 62.3%; Score 2481.5; DB 15; Length 533;
Best Local Similarity 83.3%; Pred. No. 3,4e-203;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;
Qy 203 MASGGGAPMADNEGADVGNASGNMHCSTWMDREVTITSTRTMAALPTNNHLYKQISS 262
Db 1 MATGSGAPMADNEGADVGNSSGNMHCSTWMDREVTITSTRTMAALPTNNHLYKQISS 60
Qy 263 AGTGAASNDNHGYGISTPAGYFPDNRPHCFSPRDORLLNNWNGFRPKLNFLEN 352
Db 61 QS-GASNDNHGYGISTPAGYFPDNRPHCFSPRDORLLNNWNGFRPKLNFLEN 119
Qy 323 EYVTTNGVTIANNLSTVQVPSDSSEYOLPYVLGSAHQGLPPFPADVMIPOYGY 382
Db 120 EYVTTNGVTIANNLSTVQVPSDSSEYOLPYVLGSAHQGLPPFPADVMIPOYGY 179
Qy 383 NGSOAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVPHSSVAHSQSLDRMLNPLIDY 442

Db 180 NGSOAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVPHSSVAHSQSLDRMLNPLIDY 239
Qy 443 LVIARTQOSSGAQKDLIFSRGSPAGMSVOPKXMLBPCYRQOQVSTKTKDNN 502
Db 240 LVIARTQOSSGAQKDLIFSRGSPAGMSVOPKXMLBPCYRQOQVSTKTKDNN 299
Qy 503 WTCASKYNNGRESINPGTAMASHKDEDEKFPFGSGWMI FGKESAGASNTALDNY 562
Db 300 WTCASKYNNGRESINPGTAMASHKDEDEKFPFGSGWMI FGKESAGASNTALDNY 359
Qy 563 EBEIRKATNPVATERFCTVAVNFQSSSTDPATGVDHAMGALPGWMDRVDVYLGPI 622
Db 360 EBEIRKATNPVATERFCTVAVNFQSSSTDPATGVDHAMGALPGWMDRVDVYLGPI 419
Qy 623 PHTDGHFHSPLMGSGGLKNPPQILLIKTVPANPAEBSATKFSFTTQYSTGQ 682
Db 420 PHTDGHFHSPLMGSGGLKNPPQILLIKTVPANPAEBSATKFSFTTQYSTGQ 479
Qy 683 IEMELQKENSKRNNPEVOYTSNYSKASANDFTVDNNGLYTEPRPICTRYLTRPL 736
Db 480 IEMELQKENSKRNNPEVOYTSNYSKASANDFTVDNNGLYTEPRPICTRYLTRPL 533

RESULT 5

US-10-205-942-4
Sequence 4, Application US/10205942
Publication No. US20030053990A1
GENERAL INFORMATION:
APPLICANT: University of No. US20030053990A1Ch Carolina-Chapel Hill
APPLICANT: Rabinowitz, Joseph E.
APPLICANT: Samulski, Richard J.
APPLICANT: Xiao, Weidong
TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
FILE REFERENCE: 5470-186
CURRENT APPLICATION NUMBER: US/10/205, 942
CURRENT FILING DATE: 2002-07-26
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 756
TYPE: PR
ORGANISM: Adeno-associated virus
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(2271)
OTHER INFORMATION: B19/AAV chimeric capsid coding sequence
US-10-205-942-4

Query Match 35.3%; Score 1410; DB 15; Length 756;
Best Local Similarity 42.7%; Pred. No. 1.8e-111;
Matches 312; Conservative 95; Mismatches 269; Indels 54; Gaps 16;
Qy 1 MAADGYLPMLLEBNLESEGIRBMWDLKPGAPKPRANQODDGRGLVPGYKYLGPENG 60
Db 1 MAADGYLPMLLEBNLESEGIRBMWDLKPGAPKPRANQODDGRGLVPGYKYLGPENG 60
Qy 61 KGEVPAADAALEHDKAYDOOLKAGDNPLYLRNHAADAEFOELQEDTSGGVLGRAVQ 120
Db 61 KGEVPAADAALEHDKAYDOOLKAGDNPLYLRNHAADAEFOELQEDTSGGVLGRAVQ 120
Qy 121 AKKRVLEPLGLVEGAKTAPGKRPEVQSPQESDSSSGTQGOAPAKKRLNFGOTDSE 180
Db 121 AKKRVLEPLGLVEGAKTAPGKRPEVQSPQESDSSSGTQGOAPAKKRLNFGOTDSE 180
Qy 181 SVDPQPLGEPPTAPAVGPTTASGAGAPMADNEGADVGNASGNMHCSTWMDREVT 240
Db 181 SVDPQPLGEPPTAPAVGPTTASGAGAPMADNEGADVGNASGNMHCSTWMDREVT 239
Qy 241 TSTRTMAALPTNNHLYKQISSASTGASND-----NHFGYSTPAGYFPDNRPHC 290
Db 240 TSTRTMAALPTNNHLYKQISSASTGASND-----NHFGYSTPAGYFPDNRPHC 299
Qy 291 HFSPRDMORLLNNWNGFRPKLNFLENIQVKEVT--TNDGVTIANNLSTVQVPSDS 348

```

Db 300 FFSLEFQHLEENVGSIAPDALVTITSEIAVKDVTDTGGGV-QVDTSTTGRLMLVDHE 358
Qy 349 YQLPYVLSAHGCLPPFPADVEMIPOYGYITLNN-GS0AVG-----RSSFYCLEY 398
Db 359 YKPYVYLGQCDTLAPLPIWVPPPPQAYALTVDVATGQISDPSKILASESAFYVLEH 418
Qy 399 FFSQMLRTGNFTSYFEFEEVPHSSVAHSQSLRLNPLIDLYLYLNTQNGSSAON 458
Db 419 SSFOLSTGTATMSYFPPVPPENLGCSQHFTEMNPL--YGSRLGVPDTIGDPKF 475
Qy 459 KDLFFSGSPAGMSVQPKMLPGCYROQRVSKTKTNNNSNFTWTGASKYNLNGRESII 518
Db 476 RSL-----THEDHALIQPNFPGPLVNSVSTKEGDSNTGAKALTGLSTGTSQNTRI 530
Qy 519 NPG-TAMASHKDEDEKFFPMSGVMIFGKESAGASNTALDNV-----MTDEEIKATN 570
Db 531 RFGVSOPIYHMDTKVTGTGINAISHQVTVG--NADKEYQGVGRFPNEKQQLQLO 587
Qy 571 PVATERGTVAVNFGSSSTDPATGDVHAMGALPGMWQDDRVYLGPIWAKIPIHTDHPH 630
Db 588 GLNNHTY-----FENKGTQVTDQTE-RPLMVGSVNNRRALHYESQUMSKIPMLDSEK 640
Qy 631 PS-PLMGFGILKXPPQILIKNTFVPANPPAFESATFASFTITQYSTGVSVIEML-Q 688
Db 641 TQFALGGWGLHQPPIPLK--ILPQSGPIGKISMGITTLVQYAVGIMVTMTFPLGP 698
Qy 689 KENSKRWNP 698
Db 699 RKATGRWNPQ 708

```

RESULT 6

US-09-839-996-2

Sequence 2, Application US/0983996
Publication No. US2003009010A1

GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.

TITLE OF INVENTION: Haemophilus Adherence and Penetration

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/839, 996

FILING DATE: 20-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/296, 791

FILING DATE: 25-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Tregartin, Richard F.

REGISTRATION NUMBER: 31,801

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1394 amino acids

TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-839-996-2

Query Match 3.0%; Score 119.5; DB 12; Length 1394;
Best Local Similarity 17.4%; Pred. No. 0.69;
Matches 152; Conservative 96; Mismatches 287; Indels 337; Gaps 38;

```

Qy 49 GXYTLG-----PENG--DKGEVNAADAALEHDAKYDQQLKAGNPNV----- 90
Db 217 GYSVLGDVAKAGEYBPLPIAGSKGDSGEMFYDA---EKQWLLINGLIREGNPEGE 273
Qy 91 -----LRYNDAEFOERLOEDTSFGCNLCRAVF-----QAKRYLEPLGLVEEAKTA 139
Db 274 NGQVLRKSTFDFIFERDLH--TSLYTRAANGVYTTISGNDNGGSGITQSGISEIKITL 331
Qy 140 PGKRPVEQSPQEDSSSGIGTGOQPAKKRLNFQGT--GDESEVDDPQPLGEPATPA 196
Db 332 ANMSLPL---KEKDKNHPRVDPNIIYSPRLNNGETLYEMDQKQ----- 372
Qy 197 AVQPTMASGGAPMADNNEGADV--GVASGNHCDSTW--LGDRIYITSTRTW----- 247
Db 373 --GSLIFAS-----DINQAGILYFEGNFTVSPNSQWOGAGLHVSNSVTWVKVG 423
Qy 248 -----ALPTYNNHLYKQI--SSASTG 266
Db 424 VEHDRLSKIGKGLHQAQKENGKSGISVGDGKYLEEQADDQGNKQAFSGILVSGRTV 483
Qy 267 ASNDNHVFGYSTPMGYFDFENRFCHFSPRDMORLNNNGFRRKRLNFKLENIQVKEV-T 325
Db 484 QLNDDK-----QFDIDKFY-----FGFRGRLDLNHSILFKRIQN 519
Qy 326 TNDGVTIANNLSTVQVPSDSEYQLPYVLSAHGCLPPFPADVEMIPOYGYITLNNGS 385
Db 520 TDEGAMIVNNTTQAAV-----TITGN 542
Qy 386 QAGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYA-----HSQSLRLNPLI 439
Db 543 EST-----VLPGNNINKLDYRKIAIANGMGERTDKKHGRRLILYKPTT 588
Qy 440 DQVLYLNTQNGSS-AQNKDLF-----SGSPAGMSVQPKML 479
Db 589 EDRTLLSGGTNLKGDITQKGLFPGSRPTPHAYVHLNKRWSMEGIPQGEIYVMDHDI 648
Qy 480 -----PGPCYRQORVSKT--TDNNNSNFT-----WTGAS- 507
Db 649 NRTFKAKNPOIKGSAVSVSSISGNTVSNANATFGVVDNQNTICTRSDMTGLTT 708
Qy 508 -----KYNLNGRESIINPQTA-----MASHKDEDEKFFPMS 538
Db 709 CQKVLDLDFKVINISIKTQNGSINLTDNATAVVKLAKLNGVTLTNSQ-----FTLS 763
Qy 539 -----GVMIFGKESAGASNTALD-NVMTDEEIKATNPVATERFGVAVNFGSSST 550
Db 764 NNATQIGNIRLSDNSTATVDNANLNGVHLTSAQPSLKN-----SHFSH 808
Qy 591 PATGDVHAMGALPGWV--QDRVYLGPIWAKIPIHTDGHHPSPMLMGFGELKXPPQILI 649
Db 809 QIQGDGKTVTLLENATWMPSDTTLNLLNNTLNSGAVSA-----SSNTPFRRL 862
Qy 650 KNTFVPANPPAFESAT-----KFAEITQYSTQGVSV-----EIEMLQKENS 693
Db 863 ETEFTTSAHRRNTLTVNGKSLGSGGTPOFTSLPGYKSKLXLSNDAEDGYLTSVRNTG 922
Qy 694 RANPE-----VOYTSNTAKSANVDYDNN 718
Db 923 K-EPETLEQLTLVESKDNQPLSDKLTLEND 953

```

RESULT 7

US-10-080-505-2

Sequence 2, Application US/10080505

```
Publication No. US20030073166A1
GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RT/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080,505
CURRENT FILING DATE: 2002-02-22
PRIORITY FILING DATE: 1994-10-25
PRIORITY APPLICATION NUMBER: US 09/839,996
PRIORITY FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1394
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-080-505-2

Query Match          3.0%; Score 119.5; DB 15; Length 1394;
Best Local Similarity 17.4%; Pred. No. 0.69; Indels 337; Gaps 38;
Matches 152; Conservative 96; Mismatches 287;

QY 49 GYKTLG-----PENGGL-DKGEFVNADAALEHDXADQOLKAGDNPY-----90
DB 217 GYGVLLGDVAKAGEYGLPLAGSKGDSGSPMFTYDA---EKQKWLINGLIREGNPFEGKE 273
QY 91 -----LRYNADAFQERLOEDTSFGGNLGRAYF-----QAKKVLIEPLGLVEBAGKTA 139
DB 274 NGFOLVAKSYFDEIFERDILH--TSLYTRAGNGVYTTSGNDNGGSGITQKSGIPSEIKITL 331
QY 140 PGKKRPVEGSGPOEPSSSGIGKTGQCPAKKRLNGOT---GDSESVDDPPLGEPATPA 196
DB 332 ANMSLPL---KPKDKYHNRYDGNPNYISRLNNGELTYMDQKQ-----372
QY 197 AVGPPTWASGGAPMADNNEGADGV---GNASGNWHDSTW--LGRVITTSRTW---247
DB 373 --GSLIFAS-----DINQAGGLYEGNFVTPSPNSNQWQAGIHVSNSITVTKVNG 423
QY 248 -----ALPTYNHLKYKI--SSASTG 266
DB 424 VEHDRLSKIGKGLHVOAKGENKSGISVGDGKVLLEQADQDNKQAFSEIGLVSGRGTV 483
QY 267 ASNDNHFGYSTPMGYDFENRHFCHFSRDMQRLINNMGFRPKLNFNIQVKEV-T 325
DB 484 QLWDDK--QPTDKFY-----FGRKGRDLNCHSLTFKRIQN 519
QY 326 TNDGVTTIANNLSTVOVFSDEYQLPYVLGSAHQGLPPFPADVEMIPOGYLTLINNGS 385
DB 520 TDEGAMIVNHNNTQAAV-----TITGN 542
QY 386 QAVGRSSFYCLEYFPSSQMLRTGNNFTSYTFEEVPHSSYA-----HSQSLRLNPLI 439
DB 543 EST-----VLPGNNINIKLDYRKELIAYVNGMGETDKNGHNRMLIYKPTT 588
QY 440 DQYLLVLRNTQNGSG-AONKDLF-----SRGSPAGMSVQPPKWL 479
DB 589 EDRTLLISGNTLKDITOTKGLFFSGRPTPHAYNHLNKRWSBMESIPQGEIWMDDHWI 648
QY 480 -----PGPCYRQORVSKTK---TDNNNSNT-----WTGAS- 507
DB 649 NRTFKAEINFOIKGSAVAVSRNVSIEGNMTVSNANNAITPGVVPNQWITICTRSDMTGLTT 708
QY 508 -----KTNLNGREBSITNPGTA-----MAHKODEDKFPFMS 538
DB 709 CQKVDLDTKIVINSIPTQINGSNILTDNATANYKGLAKLNGVTLNHSQ-----FTLS 763
QY 539 -----GVMIFGKESAGASNTALD-NVMTDEEIEKATNPVATERGTAIVNPOSSSTD 590
DB 764 NNAITQINILSDNSTIYVDNANLNGVHLTDSQPSLKI-----SHFSH 808
QY 591 PATGDVHAMGALPQMWV-QDRDVLQGRPWAKI PHTDGHFHPBLMGFGIKNPPPOILT 649
```

```
DB 809 QIQGDKGTVTLLENATMTWPSDTLLQNLINNSTITLNSAYSA-----SSNNTPRRRL 862
QY 650 KNTFPVAMPAPRSAT-----KEASPIITQYSGOVV-----HEMELQENSK 693
DB 863 EFTTPTSAHRRNTLTVNGKLSGQGTFOFTSSLFQYKSDKLKLSNDADGDTILSVRNNG 922
QY 694 RMNPE-----VOYTSNYAKSANVDFVDDNN 718
DB 923 K-EPETLEQTLTVESKDNOPLSDKXKFTLEND 953

RESULT 8
US-09-815-242-12713
Sequence 12713, Application US/09815242
Patent No US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haeelbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA_011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: Faeseq for Windows Version 4.0
SEQ ID NO 12713
LENGTH: 2344
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match          3.0%; Score 118; DB 10; Length 2344;
Best Local Similarity 20.4%; Pred. No. 2;
Matches 87; Conservative 57; Mismatches 182; Indels 100; Gaps 18;

QY 174 GQNGDESEVP-DQQRGEPBPAPPAVG-----PTTMSGGAPMA 212
DB 326 GNGGDDIGAFBSGVLGETGLNGAAVIGLSNAPFKLDYTHNISTPNSAKAKDPSN 385
QY 213 DNNEGADGVGNASGNMHCSTWLGDRVITTSRTWALPTYNHLKYQISSASTGAGNDN 272
DB 386 VAGGAGFAGFVTTDSGVASTYSSSTADMAKLVNQPTNN-----TFQDFDIN 434
QY 273 YFGYSTPMGYDFENRHFCHFSRDMQRLINNMGFRPKLNFNIQVKEVTTNDGVTT 332
DB 435 YNG-----DTKVMTVYAGQGTWTRNI-SPWIAKSTTNEGL-----SMTASTGAGATN 480
QY 333 IANNLSTVOVFSDSYQOLPYVLGSAHQGLP--FPADVEMIPOGYLTLINNGOAVGR 390
DB 481 LQOVQGTREYIESATQVRYVDVITGKDIIIPKTYSGANDVY-----VTIDQGSALTA 535
QY 391 SSP-----YCLEYFPS-----QMLRTGNNFTSYTFEEVPHSSYAHOSLDRIMNPL 438
```

Db 536 KGNVTSVDSASTYNDTNKTVKMTAGASVTYYFTDVKAPVTVGNOTIEVGKTMNPI 595
Qy 439 IDQVLYIANTQNSGSAQKNDLPSRSGPAGMV-QPKMPLPBPCTR--QQRSKTKTD 495
Db 596 V-----LTTTNDGTGTNT-----VTGLPGLSYDSATNSIIGTPKIGSTVAVSTD 645
Qy 496 --NNNSFTWTKSKYMLNGRESIIN-----PGTAMASHCDDEKFPMSGVMIPEGES 547
Db 646 QANNSKSTTFT-----INVDTAPVTPTPGDKSSR-VSPSPISPIATQDN 691
Qy 548 AGASNT 553
Db 692 SGNAV 697

RESULT 9

US-10-080-505-7
; Sequence 7, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DBR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-7

Query Match 2.9%; Score 117.5; DB 15; Length 1395;
Best Local Similarity 17.9%; Pred. No. 1;
Matches 125; Conservative 73; Mismatches 211; Indels 291; Gaps 31;

Qy 49 GYKTLG-----PFGNL--DKGPVNAADAALEHDKAYDQOLKAGDNPY---- 90
Db 217 GYSLVGGVVRKAGYGLPLPIAGSKDSSPMFVDA--EKQKWLINGILREGNPFSGKE 273
Qy 91 -----LRYNADAFQERLQEDTSFGNLGRAV-----QAKRVLEPLGLVEGAKTA 139
Db 274 NGFOLVRKSYDELFERDLH--TSLYTRAGGVYTTISGNDKQGSITQKSGIPSEIKITL 331
Qy 140 PGKRPVPEQSPQEPDSSSGIGKTGQPAKKTILNFGQT---GISSEVPDPQPLGEPATPA 196
Db 332 ANMSLPL--KENDKHNPNYDGPNTYSPRLNGETIYFDQKQ----- 372
Qy 197 AVGETTASGGGAPMADNEGADV--GNASGMWHDSTW--LGRVITTTSTRTW--- 247
Db 373 --GSLIFAS-----DINGAGGLYFEGNFTVSPNSQGTQAGIHSENSVTWAKNG 423
Qy 248 -----ALFTYNNHLYKQI--SSASTG 266
Db 424 VEHDRLSKIGKTLHVAQKGNKGSISVGDKVLLEQADQKQKQAFSEIGLVSGRGTV 483
Qy 267 ASNDNHFGYSTPWGVEFENRFGHGFSPDMQRLINNMGPRPKRLNFKLPIQVKEV-T 325
Db 484 QLANDK-----QPTDKFY-----FGRGRGLDLNGSLFFKXIQN 519
Qy 326 TNDVTTIANNLTSVQVFSDEXQLPYVLGSAHQGLPPPADVPMIPOYGYLLTNGS 385
Db 520 TDEGAMIVNHTTQAANY-----TI-TGN 542
Qy 386 QAVGRSSYCLEYPPSQMLRTGNFTPSYTFEYVPHSSYA-----HSQSLDLNMPLI 439
Db 543 ESI-----VLPNGNNINKLDYRKELIAYNGWFGRTDKKRNKRLNLIYKPTT 588

Qy 440 DOYLVIANTQNSGS-AQKNDLFP-----SRGSPAGMSVQPKNML 479
Db 589 EDRTLISGTLNKGITLTQKGLFFSGRPTRPANYHNLNKRSEMGISPGGELVMHDWI 648
Qy 480 -----RGPCTQROQVSKTK--TDNNNSFT-----WTGAS- 507
Db 649 NRTFKAEFQIKGSAVSVSNVSSIGNMTVSNNAATFGVBNQONTICTRSDMTGLT 708
Qy 508 -----KYNNGRESIINPETA-----NASHKDDKFPFMS 538
Db 709 QKQVLDTRKYVINSIPTQNGSINLTDNATAVKGLAKLNGVTLTNHSQ-----FTLS 763
Qy 539 -----GVMIFGKESAGASNTALD-NVMITDEEIKATN 570
Db 764 NNATQIGNIRLSDNSTATVDNANLVNGLTDSAQPSLKN 803

RESULT 10

US-09-815-242-5845
; Sequence 5845, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5845
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5845

Query Match 2.9%; Score 117; DB 10; Length 1016;

Best Local Similarity 20.3%; Pred. No. 0.7; 272; Indels 240; Gaps 40;
Matches 155; Conservative 96; Mismatches 240;

Qy 69 DAAALEHDKAYDQOLKAGDNPYLRYNHADAFOERLQEDTSFGNLGRAVFAQKKEVL-- 126
Db 1 EABAENNIENFTLK--DN-----VQSKVKIEVNNKOTAPQGVAKSEVTSNKOTIEH 54
Qy 127 EPLGLVEGAKTAPGKRPVPEQSPQEPDSSSGIGKTGQOPAKRL-----NFGQGDSE 180
Db 55 EPEVKAEIDISKEDTKEVADVAVQPKSS--VTHNAETPKVAKARSVDGSEFDITRDSK 112
Qy 181 SVDP-----PQPLGEPATPAVGPPTMASGSGAGMAD----- 213
Db 113 NVSEPTPTITQKHEFGYGSVDIQKPPDGLGVSEVTRFNVGNSGLIGALQKKNKIDF 172

```

QY 214 -----NNEGADG-----VGNASGNMHCSTWLGDRVITSTRVALPT 251
Db 173 SKDNFKRVANNHQSITGADGGLFSGGNAB-EVLTNGILGDKGLVNS--GGFKIDT 230
QY 252 ---YNNHLVQISSASGASNDNHFGYSTPWGTFDNRHCHSPDMOQLINNKGFR 308
Db 231 GYITSSMDTERKQAGG-----YRGY---GAF-----VNDSSG-- 262
QY 309 PKRLNFTLQVKEVTTNDGVTIANNLSTVQVPSDSYQ-----LPVYLGSAHQ 360
Db 263 ---NSQMGVENIDKSTN--FLANYADNSTNT---SDGFHGORLNDVLLTYVASTGKM 312
QY 361 GC-----LPPFPADVMIPOXYGILLNNGSOAVGRSSFCLEYFPQMLR 405
Db 313 RAEYAGKTWETSITDGLSKNOAVNFLTSSQRMGLNGGINANG-----MMR 359
QY 406 T---GNNFTS-----YTFEEVPHSSVSAHSQSLDRLMPLIDYLYLANTON 451
Db 360 TDJGSEFTTPAPAKTITTELEKKVEEIPRKK-----ERKFNPDLAGTERKVTNR-EG 410
QY 452 QSG-----SAON--KDLFSRGSF-AGMSVOPKMLP-----GPCYRQORVSKTKT 494
Db 411 QKEKEKITTTLKPNPLTGLVSIKGEPEKEITKDPINELTYGPEITLAFGRHDEPDKLPT 470
QY 495 DNNNSFTWTGASKYMLNGRESIINP--GTAMASHKDEDEKFPFMSGWMIFGESAGASN 552
Db 471 -----GKEEVEPGKPGIKNPEGDVVRPVDVSVTKYGPVKGDSIVEKE----- 514
QY 553 TALDNMTDEEIKATNPVA--TERPGIVAVNFQSSST-----DPATGDVHAMGALPGM 605
Db 515 -----IPPEKERRKFPDLAPGTEKTRREGOKGEKITTPTLKPNPLTGLVSIKGESKEE 567
QY 606 WQOD--RDVYLOPIMAKIPIHTDGHFHP-----SPLMGFGIXNPPPIILIKNPVPA 656
Db 568 ITKDPINELTYGPEITTPPHRD--EPDPKLPTEGKEEVPKPGIKNPEGDVVR----- 620
QY 657 NPFAEFSATKFAFITQYSTGOVSVEIEMELOKENSKRNPPEV 699
Db 621 -PVD-SVTKYGPV-----KGDSIVEKE-EIPPEKERRKFPDL 655

RESULT 11
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match      2.9%; Score 116.5; DB 10; Length 5795;
Best Local Similarity 19.1%; Pred. No. 11;
Matches 128; Conservative 86; Mismatches 250; Indels 207; Gaps 35;

QY 177 GSESVDPDPLCEPATPAAGPTTMAAGGAGPMAADNEGADGVNAGSNMHCSTWLG 236
Db 1696 GNSGSAADYFKLSNGSAIDA---TITWVGQAPNDONRIGDI--NVTANILID---G 1747
QY 237 DRVITTSITWALPTYNHLYKQISSASGASNDNHFGYSTPWGTFDNRHCH- 291
Db 1748 ETPPIKTKAT-----YKVS-----SVPKHFETNRCGAVFPQVSD 1782
QY 292 -FSPRDMOQLINNMGFRPRKRLNFKLENI-----QVKEVTTNDGVTIANNLSTVOVS 345
Db 1783 VIDAKQYKPVNDSTQMPQRMNFOFTNSYGBSKDVVGISTRIDRYVDNHOTIILILA 1842
QY 346 DSEYQLPYVLGSA-----HQCPLPPADVMIPOXYGILLNNGSOA 387
Db 1843 KYKPDPRIDGNSVYTKAGLTNQIKINVLSSSITKFKAD--NTP---LTIITTYG 1896
QY 388 VGRSSFYCL-EYFPQMLRTGNNFTFSYTFEEVPHSSVSAHSQSLDRLMPLIDYLYL 446
Db 1897 SGNATAVVTSDALPENGVIKARS---SITMNVYTTQDEHGRAIDVTNRESVD----- 1946
QY 447 NRQONSQS-----AOKDLFSRGS-----PAGMSV-----OPKMN- 478
Db 1947 ---SNDSATVYTPQLOATTEGAVFTKGDDGFDFGVERFIQNPFGATVAMHNDPDWK 2003
QY 479 -----LPGCYRQORVSKTKTDNNNSFTWTGASKYMLNGR 514
Db 2004 NTYGNHTKATAVVTLPBGGGSTRNVEYVKYYPANAKAPRDKGQULT-----NGT 2054
QY 515 EST---INPGT-----AMASHKDEDEKFPFMSGWMIFGESAGASNTALDNMT-- 561
Db 2055 DAINYITFPDNTYTNQITTAAMANRQOPNNQ---QAGVQHLNVDTYPTGITAARVPTVN 2111
QY 562 -DEEIKATNPVATERFGVAVNFQSSSTDPATGDVHAMGA--LP--GNV--WQDDVYL 614
Db 2112 VYQFEFPQTSYTTVG-GTLANGTQ-----ASGVAHMONANGLPDGTYYKNNNAATGT 2164
QY 615 QGPIMAKI--PHT-----DGHFSPBLMGFGIKN--PPQILIKNTPVPA--- 656
Db 2165 NDANMAAMKRNPAKYNKAVIVYINGHTFATSLPAKFYKDVQPAKPTVTEAAGATTI 2224
QY 657 NPFAEFSATKFAFITQYSTGOVSVEIEMELOKENSKRNPPEVQYTSNYAKA 709
Db 2225 TEGANQTVHTAGNVTTADKLVIKRNGV---VTFTRRNNTSPVKE---ASAATVA 2277
QY 710 NVDFTYDNGGL 720
Db 2278 GIAGT--NNGI 2286

RESULT 12
US-10-022-809-2
; Sequence 2, Application US/10022809
; Publication No. US20030113407A1
; GENERAL INFORMATION:
; APPLICANT: LIU, Yi-Shin
; APPLICANT: LIU, Chang-Hsueh
; APPLICANT: CHU, Wen-Shen
; TITLE OF INVENTION: TRANSGUTAMINASE GENE OF STREPTOCOCCUS LACTICUS AND THE
; FILE REFERENCE: U-013779-2
; CURRENT APPLICATION NUMBER: US/10/022,809

```

CURRENT FILING DATE: 2002-03-29
PRIOR APPLICATION NUMBER: US 10/021,678
PRIOR FILING DATE: 2001-12-12
NUMBER OF SEQ ID NOS: 3
SEQ ID NO 2
LENGTH: 410
TYPE: PRT
ORGANISM: Streptococcus lactococcus
US-10-022-809-2

Query Match 2.9%; Score 115.5; DB 15; Length 410;
Best Local Similarity 22.0%; Pred. No. 0.24;
Matches 76; Conservative 43; Mismatches 151; Indels 75; Gaps 20;

QY 4 DGYLPDM--LEDNISEGIREW--WDKPGAPKPKAKQOQDORGLVPGYKTLGPENG 58
DB 97 DVPKPSGRATYNNYIRKQOVYSHRDKQOQTEORWLSYGCYVTWVNSGQY-- 154
QY 59 LDKGEPVNAADAALHDKAVDQOLKAGDNFYLRYNHADAFCERLOEDT---SFGNLTG 115
DB 155 -----PNNRLAFAPFDEDK-YKMLKNG-RP--RSGTRAFEGBRVAKDSDEAKGQRA 205
QY 116 RAFAOAKKRYLEPLGLVEGAKTAPGKKR-----PVEQSPQEPDSS 156
DB 206 RDVASVWAKALE--NAHDEGAYLDNLKELANGDARNEDARSPFYSALRNTSPFDNRN 263
QY 157 SGIGKTOQOAPKRL-----NFGOTGSESVDPQPLGEPRA--TPAAVGPPTWASGGA 209
DB 264 GG---NHDPKMAVYISKHFGSGQDRSGSSDKKYGDPAPFPDRGTGLVMSRRPNT 319
QY 210 PMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRT-WALPTVNNHLYKOISSASTGAS 268
DB 320 PRSTSPGESFV-----NFDYGMFAGQTEADADKTYW---TGNHYH-----APNGSL 364
QY 269 NDNH-YFGYSTPW--GFEDNR--FHCHSPRDMQRL---INNWW 305
DB 365 GAMHVESKFRNMGDGYSDPDRGAYVTVFVKSWNTAPDKVTCQM 409

RESULT 13
US-10-170-682-3
Sequence 3, Application US/10170682
Publication No. US20030104975A1
GENERAL INFORMATION:
APPLICANT: Johan AUWERX
APPLICANT: Pierre CHAMRON
TITLE OF INVENTION: Cofactor-based screening method for nuclear receptor
FILE REFERENCE: 017753-163
CURRENT APPLICATION NUMBER: US/10/170,682
CURRENT FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 60/297,772
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 1441
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: SRC-1. Cloning of full-length human steroid receptor
OTHER INFORMATION: coactivator-one
US-10-170-682-3

Query Match 2.8%; Score 112; DB 15; Length 1441;
Best Local Similarity 19.3%; Pred. No. 3.2;
Matches 140; Conservative 92; Mismatches 282; Indels 210; Gaps 32;
QY 48 PGYKTLGPNLDDGEPVNAADAALHDKAVDQOLKAGDNFYLRYNHADAFCERL--Q 105
DB 326 PSYRFI-----LNDGTWLSAHTCKCLCYPSQPMQ-----FTMGHIIIDREHSGLSPO 374

QY 106 EDTSPGKNGRAVFAKKEVLEPLGLVEGAKTAPGKKRVE-----QSPQEPD 154
DB 375 DTNNGMSIPR-VNPSVNPISPAHGVARSSTLPENSNMNVSTRINRQSSDLHSSSHN 433
QY 155 SSSGIGKTOQOQ-----AKRLNFGOTGSESVDPQPLGEPPTAAV-----GPTM 203
DB 434 SSNSQSGFQSGPSQIVANVALNKQASQSSKPSLN-LNNPMEGTGSLAQFMSPRQ 492
QY 204 ASGGA--PMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTVNNHLYKOIS 261
DB 493 VTGILATRPMPNNSPPNIST-----LSSPVGNTSAC-----NNNRKSNIP 537
QY 262 SASTGASND--NHFGYSTPWGYFDNRFRCHSPRDMQRLINNNGFAPKRLNFKLFINI 319
DB 538 VTSLOGNNEGPNNNSVFSAS-----SPYLKQWSSQNS-----PSRL-----NI 575
QY 320 QYKEVTTNDGVTIANNLSTVQVPSDSEYQPLVYLGSAHQGLPPFPADVFIPOGYVL 379
DB 576 QPAAKASKDN-KEIASLTNEMIQSDNSSDGKPLDSGLH----- 614
QY 380 TLNNGSOAVGRSFGCYCLEYFSPQMLRTGNNTFSTYFEEVFFHSSYAHQSOLDRLNMPLI 439
DB 615 --NNDRLSDGDSKYQSOTSHLVQLTT-----TAQQLRHA-----DITSCKDVL 658
QY 440 DQYLYYIANTQNSG-----SAQNKDL--LFSRSPAG--MSVQPKMLPGPCY 484
DB 659 SCTGTSNSASANSQSGSCPSHSLTARHKLIRLLAQEGSPDITTLISVEP----- 709
QY 485 RQQRVSKTIDNNNSFTWTGASKYLNNGRESIINQCTAMASHKDEDEFFPMSVMI-- 542
DB 710 -----DKKDSASTSVSYTG-----OVQNSST--KLEIDASKKESKDHQRLRYLIDK 755
QY 543 FGKESAGASNTALDNMTDEEIKI-----TNPATERFCTVAVNFQSSSTDPAQD--- 595
DB 756 DEKDLSTPLSDVDKVKYKEKQMDPCNTNTPMTKTPPEIKIKLDAQSQFTADLDQFD 815
QY 596 -----VHANGALPGWQDRDVTYLGPIWAKIPIHTDGHFHPSPGLMGFLKAPPPOLIK 650
DB 816 QLLPLEKAQQLPGLCETDR--MDGAV-----TSVITIK 846
QY 651 NTPVPANPPAFESATTFASITTOYSTGOVSVEIEMLQKENSRRNP-----EVOYTSNTA 706
DB 847 SETLPAS-----LOSATARPSTRNLRLPELELAIIDNOFGQGTGQDLPMTNTV 896
QY 707 KSN 710
DB 897 TAIN 900

RESULT 14
US-09-862-027-80
Sequence 80, Application US/09862027
Patent No. US20020142428A1
GENERAL INFORMATION:
APPLICANT: Hodge, Martin R.
TITLE OF INVENTION: No. US20020142428A1 kinases and uses thereof
FILE REFERENCE: 35800/234862
CURRENT APPLICATION NUMBER: US/09/862,027
CURRENT FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 09/345,473
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 82
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 80
LENGTH: 1265
TYPE: PRT
ORGANISM: Homo sapiens
US-09-862-027-80

Query Match 2.8%; Score 111.5; DB 11; Length 1265;
Best Local Similarity 24.9%; Pred. No. 2.9; 70; Indels 75; Gaps 12;
Matches 57; Conservative 27; Mismatches

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 26.3842 Seconds*

(without alignments)
2681.720 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989

Sequence: 1 MADAGYLPDWLEDNLESGIR.....NNGLYTEPRPIGRVLTPL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:.*
2: p1r1:.*
3: p1r2:.*
4: p1r3:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	54.6	504	1 VCPV3A	coat protein - ade
2	2168.5	54.4	732	2 SS2210	coat protein VP1 -
3	798.5	20.0	673	1 VCPV35	coat protein VP1 -
4	600.5	15.1	781	1 VCPV19	coat protein VP1 -
5	482	12.1	729	1 VCPVNA	coat protein VP1 -
6	444.5	11.1	729	1 A60006	coat protein VP1 -
7	444.5	11.1	727	1 VCPV1F	coat protein VP1 -
8	439.5	11.0	722	1 VCPVME	coat protein VP1 -
9	431.5	10.8	727	1 VCPVPP	coat protein VP1 -
10	426	10.7	723	1 VCPVPP	coat protein VP1 -
11	424.5	10.6	718	1 VCPVPM	coat protein VP1 -
12	421.5	10.6	748	1 VCPVCP	coat protein VP1 -
13	407.5	10.2	737	1 VCPVCD	coat protein VP1 -
14	404	10.1	722	1 VCPV12	coat protein VP1 -
15	385.5	9.7	722	1 VCPVCN	coat protein VP1 -
16	377	9.5	716	1 VCPV2M	coat protein VP1 -
17	374	9.5	587	1 B44276	coat protein VP1 -
18	199.5	5.0	810	2 A44054	offl protein - Jun
19	198	5.0	584	2 S49594	capsid protein VP2
20	191	4.8	702	1 VCPVAP	coat protein VP1 -
21	143.5	3.6	648	2 S50856	whn protein - rat
22	139	3.5	931	2 T49710	related to glucan
23	131.5	3.3	1446	2 A18344	immediate-early pr
24	130.5	3.3	1145	2 T18235	transcription acti
25	130	3.3	1338	2 A86827	MAP kinase kinase
26	129	3.2	1072	2 A86827	hypothetical prote
27	124.5	3.1	2493	2 A55481	adenylate cyclase
28	124	3.1	1296	2 C82521	hemolysin-type cal
29	124	3.1	3078	2 T28432	variant-specific s

30	123.5	3.1	690	2 AB0124	probable TonB-depe
31	123.5	3.1	2529	2 B64635	toxin-like outer m
32	123	3.1	1142	2 T37455	enamelin precursor
33	122.5	3.1	880	1 SYBSVS	valine-tRNA ligase
34	121	3.0	635	2 P96660	protein P2K1.10
35	121	3.0	1113	2 S28925	nuclear pore compl
36	121	3.0	1549	2 I48103	type VII collagen
37	120	3.0	1379	2 T37752	hypothetical serin
38	119.5	3.0	1180	2 E86719	hypothetical prote
39	119.5	3.0	1394	2 S60762	IGA-specific berin
40	119.5	3.0	2004	2 AC0314	probable membrane
41	119	3.0	667	2 A41311	transcription fact
42	118.5	3.0	2399	2 H71879	toxin-like outer m
43	118	3.0	739	2 T52289	probable transkeo
44	118	3.0	956	2 T08144	myosinase-binding
45	117.5	2.9	1076	1 A35622	nuclear pore prote

ALIGNMENTS

RESULT 1

VCPV3A coat protein - adeno-associated virus type 2

C/Species: adeno-associated virus type 2

C/Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jul-1999

C/Accession: A03698

R./Srivastava, A.; Lusby, E.W.; Berns, K.I.

J. Virol. 45, 555-564, 1983

A./Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.

A./Reference number: A03694; MUID:83164299; PMID:6300419

A./Accession: A03698

A./Status: translation not shown

A./Molecule type: DNA

A./Residues: 1-504 <SRI>

A./Cross-references: EMBL:J01901; NID:G209616; PIDN:AAA42376.1; PID:G209621; EMBL:M12405;

A./Superfamily: adeno-associated virus coat protein

C/Keywords: coat protein

Query Match	54.6%	Score 2177.5;	DB 1;	Length 504;
Best Local Similarity	82.5%;	Pred. No. 1.2e-134;		
Matches 392;	Conservative 33;	Mismatches 47;	Indels 3;	Gaps 2;
QY	203	MASGGAPMADNNEGADVGNSGNWCHDSTWLGDRVITTSPTTMTLPTNNHLYQOIS	262	
DB	1	MATGGAPMADNNEGADVGNSGNWCHDSTWLGDRVITTSPTTMTLPTNNHLYQOIS	60	
QY	263	ASTGASNDNHYFGYSTPMGYPDFNRPHCHFSPPDMQRLINNMGFPKRLNPLFNIOVK	322	
DB	61	QS-GASNDNHYFGYSTPMGYPDFNRPHCHFSPPDMQRLINNMGFPKRLNPLFNIOVK	119	
QY	323	EVTNDGVTTIANNLSTVQVPSDSYQLPYVIGSAHQGLPPFPADVFMIPOYGLTLN	382	
DB	120	EVTQNDGVTTIANNLSTVQVPSDSYQLPYVIGSAHQGLPPFPADVFMIPOYGLTLN	179	
QY	383	NSGQAVGRSSFYCLEFSPQMLRTGNPFSTYEEVPPHSSVAHSOSIDRLMNPIDQY	442	
DB	180	NSGQAVGRSSFYCLEFSPQMLRTGNPFSTYEEVPPHSSVAHSOSIDRLMNPIDQY	239	
QY	443	LYVINRTONQSGSAQKDLFFSRGSPAGMSVOPKMLPGPCYRQCRVSKTKTDNNNSNFT	502	
DB	240	LYVLSKRTNTPSGTTTQSRLOFSQAGASDIRDSRMLPQPCYRQCRVSKTKTDNNNSNFT	299	
QY	503	WTGASRTNNGRSGTINPGTMAASHKODDKFPFMGVMIFGKESAGASNTALDNTVITD	562	
DB	300	WTGATKYHLNGRSLVNP--AMASHKODEKFPFMGVMIFGKESAGASNTALDNTVITD	357	
QY	563	EEBIRKATNPVATREFTGAVNPPSSSTDTATGVHMGALPGVWQDRTVYLOGPTWAKI	622	
DB	358	EEBIRKATNPVATREFTGAVNPPSSSTDTATGVHMGALPGVWQDRTVYLOGPTWAKI	417	
QY	623	PHTDGHFSPMLGFGFLKNPQPOLIKNTPPANPAPASATKFPSTFOYSTG	677	

Db 418 PHTDGHFPRPLMGFGGLKHPRPOLIKNTVPVNPBPSTTFSAKXASFTYQSTG 472

RESULT 2

S52210

coat protein VP1 - muscovy duck parvovirus

N:Alternate names: VP1 protein

C:Species: muscovy duck parvovirus

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-2000

C:Accession: S52210

R:Zadori, Z., Erdel, J., Nagy, J., Kisari, J.
submitted to the EMBL Data Library, September 1993

A:Reference number: S52209

A:Accession: S52210

A:Molecule type: DNA

A:Residues: 1-732 <ZAD>

A:Cross-references: EMBL:X75093, NID:g609091, PIDN:CAAS2384.1, PID:g609093

A:Experimental source: strain FM

C:Genetics:

A:Gene: VP1

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

Query Match	Similarity	54.4%	Score 2168.5	DB 2	Length 732
Best Local	Similarity	55.2%	Pred. No. 8.1e-134		
Matches	418	Conservative	90	Mismatches 186	Indels 63
Qy	9	DWLEDNLSSEGIKREWMWDLKPGAKPKRANQKQ-----	DDGRGLVLPQYKYLGPNN	57	
Dh	10	DWYE-----TAASWMLHAKAGAPKPKSNQOOSQSVSTDRKPKQRKDNRRGFVLPGYKVVGPBN		65	
Qy	58	GLDKGEPPNADAALAEHDKAYVDQOLKAGDNLYLTYNHADAERFORLEEDTSFGGNLGRA		117	
Dh	66	GLDKGEPPNKAADSVALEHDKAYVDQOLKAGDNLYLTYNHADAERFORLEEDTSFGGNLGRA		125	
Qy	118	VFOAKKRVLEPLGLVEBEGAKTAPGKKRPVYESQSPQESDSSSGIGKTGOQPAKKRLNFGQNG		177	
Dh	126	VFOAKKRVLEPLGLVEBEPVNTAPAKK-----SSG-KLTDHPYVKKPKLSE--		170	
Qy	178	DSESVDPDQPLG-----EPAPTPAAVGPPTMAAGGCAPMADNNEGADGVGNASG		226	
Dh	171	--ENSPSPSNSGGEASAAATBGESEVAAP-----NMEEGSGAAGDSAGAGDGVGNASG		222	
Qy	227	NMHCDSTWLGRIYVITTSRTYMLPTVYNHLYKQISSASTGASNDNHYFGYSTPMGYFPBN		286	
Dh	223	NMHCDSQWLGRIYVITKTRTWLPSYNNHMYAITSQNPDSN-IQYAGYSTPMGYFPBN		281	
Qy	287	RPHCFSPRDMQRLINNMGFRPKKLNFELNIQVEVTTNDGVTTIANLITSTVQVPSD		346	
Dh	282	RPHCFSPRDMQRLINNHWGIRPKALKFKFIENVQVEVTTQQTQKTIANNLTSTIQIFTD		341	
Qy	347	SEYOLPYVLGSAHQCLRPBPADAVMIQOYGLTIN--NSGQAVGRSSFYCLEYFPSPQM		403	
Dh	342	NEHOLPYVLGSAITBGTMPPPPSDVVALPYGYCTNMHNSGAFNDRSAFYCLEYFPSPQM		401	
Qy	404	LRTGNPFYFSTFEFVPHSSVAHSQSILDRMLNPIDQYLYLNRTONOSGSAQNKDLTF		463	
Dh	402	LRTGNPFSEFSEFVPHSHMFHSQDDLRLNPLDQYLMHFSEV-NGGRNAQ-----F		455	
Qy	464	SRGSDAGMSVOPKMWLPGBCTRQORVSKTK--TDNNNSNFTTGASKYNLNGRBSIINBG		521	
Dh	456	KKAAYKGAFGAAGRMNLPGPKLIDQCVRAVSGGTDVYANVMSINSKNKPYLKDREYLTLPQ		515	
Qy	522	TAMASHKDEDEKPFPMSSGVMIFGKE--SAGASNTALDNVMTIDESEIKATNVAVATERGT		579	
Dh	516	PVATTHIEDQSSVDAQNIIGIAKQPRYSGSTAGISIDMTVDEGSLAFTNVGKRPYGL		575	
Qy	580	VAVNFQSSSTPATGDVAMGALPGMWQDRDVLVLOGPIWAXIPPTDGHFHSPLMGGRG		639	
Dh	576	TVTNEQNTTTAPTAAEVLGALPGMWQNRDIYLGPIWAXIPPTDGHFHSPLMGGRG		635	
Qy	640	LKNPPOLLKNTPLVPANPAPAFSKTASFTIQTQSVIEIMELQENSKAMNBEV		699	

D6 636 LHHPPQYVITKTPPRADPEPLEYVNQKNKSIIITQYSIGQCVEHWMLREKNSKRNNPEI 695

D7 700 QYTSNAXASANDFTVDNNGLYTERDPIGRTYLTRPL 736
|::||: : | : |||||::|
D8 696 QFTSNCGNRITSIMFAPNETGGVDEBRLLGRFYLYQLNL 732 .

RESULT 3
VCPVB5
coat protein VP1 - bovine parvovirus
N:Contains: coat protein VP2
C:Species: bovine parvovirus
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #ext_change 16-Jul-1999
C:Accession: A26104
R:Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.
J. Virol. 60, 1085-1097, 1986
A>Title: Complete nucleotide sequence and genome organization of bovine parvovirus
A:Reference number: A26104; PMID:87061184; PMID:378314
A:Accession: A26104
A:Molecule type: DNA
A:Residues: 1-673 <GENE>
A:Cross-references: EMBL:M14363; NID:G9333454; PIDR:AMB59847.1; PID:sg60805
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
E:138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match	20.0%	Score 798.5	DB 1	Length 673
Best Local Similarity	28.6%	Pred. No. 2.4e-44		
Matches 214	Conservative 105	Mismatches 275	Indels 155	Gaps 21
Qy	30	PKPKANOQKDDGGRLVPGYKYLGPFGNLDKGEPPVNAADAALAEHDKAYDQOLKAGDNP	89	
Db	3	PTNKANSEK-----GLTLPNGNYLGPFNSTLPAGAPVVKADAAAKRHDGVDLLKEGKNP	57	
Qy	90	YLRYNHADAERQERLOEDTSGGMLGSAVPQAKKRYLEPLGLVEGAKTARQKRPVEOS	149	
Db	58	YLRYNHDDQNLIDELKDDTSGGKLANGVFIKALPAL-----PQTSKGDBALKRK	111	
Qy	150	POEDSSSGIGKTOQOQPAKKRLNFGQTDGDSBVDPOPLGEP--PATPAVAGPTTMASSG	207	
Db	112	LYFARSKGAKKARREBPASTSNQONNEVSNDIENDEAGNQPIELATRSVVGSGVGSGGG	171	
Qy	208	GAPAADNNEGADGVGNASGNMHCSTYLGDREVITTSRTMALPYNNHLYXQISASRGA	267	
Db	172	-----RGGSGVYSTGGTGGTGTISENVIVTKNTKQICDINGKGLYAS-EVYLNIGD	222	
Qy	268	SNDNHYEGYSTPMGYPDPNRPCHPFSRPMQRLINNMGFPRKRLNFXLIVQKEVTTN	327	
Db	223	TAHQY-AITTPMVFYFNQYSSHSFENDQHLVNDYERFPRKMIYAVNLQIKQIMTD	281	
Qy	328	DGVTTIANN-LTSYVOYFSDSEYOLPYVLSAHOGLCPRPFADYFMIOYGYLTL-----	381	
Db	282	GAMGTAVNNDDLTAGMHLFCGDHRYPYVQHEPMDQCMELRNSIWELRYQYVITPAISV	341	
Qy	382	-----NNGSQAVGRS-----SFYCLEYPPSQMLRTGNNFTFSYTFEEVYFHSVYASOSLDRM	435	
Db	342	DNNTTNVVEHLKGVPLMYLHNSDHYLRNG-----RIV	376	
Qy	436	NPLIDQVLYLYNFTQ-----QSGSAQNKULLSRSGP-----AGMSVQPKMLP	480	
Db	377	RIYIOLWRLEMRDKQHHIOMASDDVOSTGQOKNILLIORTQPNKORFQNALRTRSNMS	436	
Qy	481	GPCRQGRVSKTKDNNNSNFTMGASKYMLNGRESIINPTAMASHDEDDKFPMSGV	540	
Db	437	GP-----GIARGTNATLQTOAGALYTMVINGAD-----VSGV	470	
Qy	541	M-----IFGKE-----SAGASNTALDNYMITDEEIKATNPVATE	575	
Db	471	RAVAVGSTPIYGGQPESDILRLRYASAAEQGNPILN-----AARH	516	
Qy	576	RFGVYAVNPOSSSTDPATGVDHMAAGALPGMWQDRDVLVLOGIYAKIPIHTGHPHPSLM	635	
Db	517	TFTRBATKLTGSGADGDYKEMWMLPNDQIMDSAPISRYNPIYWKVPRVNRKLTLLDPO	576	

QY 636 GGFGLKNPPQILIK-NTVPANPAEFSAITKFSITQSGVSEIEMELQKNSK 693
 Db 577 GSIMSHRPRTIFIKLARIPIVPGND-----SPLNIYVTOQVCEVMEVEKGIK 627
 QY 694 RMNEVOYTSNYSKASAND-FTVDNNGLY 721
 Db 628 NMREYVMS---ATNMSVDAYTINNAGY 653

RESULT 4

VPV19
 coat protein VP1 - parvovirus B19 (strain Au)

C/Species: parvovirus B19
 A/Note: host Homo sapiens (man)

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
 C/Accession: A24299
 R/Shade, R.O.; Blundell, M.C.; Colmore, S.F.; Tattersall, P.; Astell, C.R.

J. Virol. 58, 921-936, 1986
 A/Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated fr

A/Reference number: A24299; MUID:86200451; PMID:3701931

A/Accession: A24299

A/Molecule type: DNA

A/Residues: 1-781 <SHA>

A/Cross-references: EMBL:M13178; NID:G333375; PIDN:AAA6867.1; PID:G333377

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein

Query Match 15.1%; Score 600.5; DB 1; Length 781;
 Best Local Similarity 26.6%; Pred. No. 2.6e-31;
 Matches 182; Conservative 95; Mismatches 300; Indels 107; Gaps 20;

QY 47 LPGYKYLGFPGNGLDKGEFVNADAALHEDKAYDQOLKAGDNPYLKRNHDAEFQELQ 106
 Db 125 LFGNYVPGNEHLOAGPPQSAVDAARHDFRYQLAKLGINPYTHVTADIELKNIKN 184
 QY 107 DTSGFGLGAVFOAKKRVLEPLGLVEGAKTALGKRPVQSPQEPDSSGIGTKQOP 166
 Db 185 ETGQAQVVDYF-----TLKGAAPVAHQ-----GSLPEVP 217
 QY 167 AKKRLNFGQGTDSSEVDPQPLGEPATPAVGPPTMASGGGADNNEGADGVNAGS 226
 Db 218 AYNA-----SEKPSMTSVNSAEASTGA-----GGG-----GSNSVKS 250
 QY 227 NMHCDSTWLDGRVTTSTRFALPTNNHLYKOISSASTGASND-----NHFGY 276
 Db 251 MWSEGAITFSAISVCTFSRQFLIPYDEHHYKVFSPASSCHNASGKEAKVCTISPIMGY 310
 QY 277 STPGYFDFNRFHGFSPRDMQRLINNMWGFPRKLNFKLPIQVKEVT--TNDGVTTIA 334
 Db 311 STPRVYDFNALNLFPSPLEFQHLIENYGAIPALVTYTSIELAVKQVTDKTKGGV- QVT 369
 QY 335 NMLTSTVQVPSDSEYQPLPYLGSAAHOGCLPEPADVEMIPQYGLTLNN- GSQAVG- 389
 Db 370 DSTGRCLMLVDHEKYPYVLGQGDPLAPELPIWYVFPQYAVLVGADVNTQGISDSK 429
 QY 390 -----RSSFCLKTFPQOMLRTGNNFTPSYTFEEVPHSHSVAHQSGLRLMPLIDOLY 444
 Db 430 KLAEESEAFYVLEHSSPOLGTGTASMSYKFPVPPEPNELEGCQHYEMWNP- YGS 486
 QY 445 YLNRTOQSGAOKKDLIFSRGSPAGMSVOPKMLPGPCYRQGVSTKTKDNNNSNFTWT 504
 Db 487 RLGPVDTLGGDPKRS- ----THEDHAIQPNMGPBLVNSVSTKEGDSSTGAGALT 541
 QY 505 GASKYINLNGRESITNPG-TAMASHKDDDEKFFPMGVMIFGKESAGASNTALDNV----- 558
 Db 542 GLSTGTSQNTIRISLRPGPVSQPYAHMDTKVYTGINAIISHQTYG---NAEDKEYQGV 598
 QY 559 -MITDEEIKATNPVATERGTVAVNFOSSSTPBAIGDVHAMALGGMWQDDVYLOG 616
 Db 599 GRPNEKQOLQGLNNHTY-----FPNGTQOQYTDLE-RPLWGSVNNRRALHYES 651
 QY 617 PIMAKIHTDGHFHS-PLMGFGGLKNPPQILIKNTVPANPAEFSAITKFSITQYS 675

Db 652 QLMSKIPNLDSEFKTQPAALGGMGLQPPQIFLK--ILPQSPIGSIXSMGTLTVQVA 709
 QY 676 TGOVSIEIEMEL-QKNSKXWME 698
 Db 710 VGIMTVMTPEKLGPRKATGRMNPQ 733

RESULT 5

VPVNA

coat protein VP1 - porcine parvovirus (strain NADL-2)

N/Contains: coat protein VP2
 C/Species: porcine parvovirus

C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
 C/Accession: B33743; D48472
 R/Vasudevacharya, U.; Basak, S.; Srinivas, R.V.; Compans, R.W.

Virol. 173, 368-377, 1989
 A/Title: Nucleotide sequence analysis of the capsid genes and the right-hand terminal pa

A/Reference number: A33743; MUID:90085785; PMID:2596019

A/Accession: B33743

A/Molecule type: DNA

A/Residues: 1-729 <VAS>

A/Cross-references: GB:M32787; NID:G332983; PIDN:AAA46917.1; PID:G332985

R/Bergeron, U.; Menezes, J.; Tjissen, P.

Virol. 197, 86-98, 1993

A/Title: Genomic organization and mapping of transcription and translation products of t

A/Reference number: A48472; MUID:94025614; PMID:8212558

A/Accession: D48472

A/Molecule type: DNA

A/Residues: 11-729 <BER>

A/Experimental source: NADL-2, ATCC VR-742

A/Note: sequence extracted from NCBI backbone (NCBIN:138789, NCBIP:138794)

C/Genetics:

A/Intons: 10/1

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein; glycoprotein

F:151-729/Product: coat protein VP2 #status predicted <VP2>

F:33,112,198,282,330,433,471,573,604,651/Binding site: carbohydrate (asn) (covalent) #str

Query Match 12.1%; Score 482; DB 1; Length 729;

Best Local Similarity 25.5%; Pred. No. 1.3e-23;

Matches 197; Conservative 111; Mismatches 310; Indels 154; Gaps 35;

QY 43 RGLVLPYKYLGFPGNGLDKGEFVNADAALHEDKAYDQOLKAGDNPYLKRNHDAEFQ 102
 Db 9 RGLVLPYKYLGFPGNGLDKGEFVNADAALHEDKAYDQOLKAGDNPYLKRNHDAEFQ 68
 QY 103 RLQEDTSPGNTLGRAVFOAKKRVLEPLGLVEGAKTALGKRPVQSPQEPDSSGIGK- 161
 Db 69 ETHEADYGGKIGHYFFRAKRAAPKL-----SEFDSPTSQPVRYSRPNHPSKPRGKR 125
 QY 162 -----TGQOPAKKRLNFGQGTDSSEVDPQPLGEPATPAVGPPTMASGGGAPMA 212
 Db 126 PABRHIFINLAKKAKGTSTNNSMSSENVQNPINAGTEL-SATNGESGGGGGG- 181
 QY 213 DNEGADGVNAGS--NMHCDSTWLDG--RVTTSTRFALPTNNHLYKOISS--SAST 265
 Db 182 -GGRGAGGVSVSTGTNNQTEFQYLDGLVRIYAHASRLIHLMPHEHYTKRIHVLNSES 240
 QY 266 GAS-----NDNHFGYSTPMGYPFNRFHGFSPRDMQRLINNMWGFPRKLNFKLPIQV 321
 Db 241 GVAGQVQDDAHQMTTPMSLIDANAMGWENADQOLISNNTETINLVSFEDEIRNVVL 300
 QY 322 KEVT--TNDGVTTIANLSTVQVPSDSEYQPLPYLGSAAHOGCLPEPADVEMIPQY 378
 Db 301 KITTESATSPRTKIYNNDDLASIMVALDITNNLTLPYPAARSETLGFYPLPTKPYQY 360
 QY 379 L-----TLNMSQAVGRS-----SFYCLE-YFPQOMLRTGNNF--TFGYTPE 417
 Db 361 YLSCIRNLNPPYTGGSQOITDSIQGLHSDIWFYITEANVPHILRTGDEFTGTYHHD 420
 QY 418 EEPFSSVYHSSQSLDR-----LNNPLI--DOVLYLNRTOQSGAOKKDLIFSRGS 467

Db 421 TKPL--KLTHSMQTNRSGLPPLKLTPTTEGQHPGLTLPANTRKGYHTINNSYTEAT 478
 Qy 468 PAGMSVQPKNMLPQPCYRQQRVSKTKTDNNNSNFTWGASKYINANGRESIINP--GTAMA 525
 Db 479 ----AIRP-----AAGVNTPYNMFYSNGSGPF-----LTPIVPTADT 512
 Qy 526 SHKODEDKFFPMGSMWIFGKESAGASNTALD---NVMITDEEIK--ATNP-----VAT 574
 Db 513 QYNDDEP-----GAIPTMYOYGHGLTTSQELERTFENPQSKCGAPK 557
 Qy 575 ERFETVA--VNFQSSS-----TDPATG--DVHMGALP-----GMWQODRDVYLQGPIMA 620
 Db 558 QQFNOQAPLNENTNNGTLTSPDPIGKSNMFMNTLTNTYGFALTALNNTAVFPNGQIMD 617
 Qy 621 KIPHTD--GHPHSPLMGFGIK--NPPQILIKNTPPANPAEFA--TKRASFITQYST 676
 Db 618 KELDTDKPRHL---VTAPEVCNNPPQGLFVKIAP---NLTDENADSPOQPRITITSN 671
 Qy 677 GQVSEIEEMELQKENSKRMPNPEVOYTSNVAKSANVDFVNDNGLYTEPRPIG 728
 Db 672 FWMKGTLTFTAKMRSNMNMPIOQHTT-----TAENIGNYI--PTNIG 712

RESULT 6

A6006 coat protein VP1 - porcine parvovirus (strain 90HS)

N/Contains: coat protein VP2
 C/Species: porcine parvovirus
 C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
 C/Accession: A6006
 R/Sakurai, M.; Nishimori, T.; Ushimi, C.; Nakajima, H.
 Virus Res. 13, 79-86, 1989
 A/Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.
 A/Reference number: A6006; M01D:89319168; PMID:2750278
 A/Accession: A6006
 A/Molecule type: DNA
 A/Residues: 1-729 (SAK)
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein; glycoprotein
 F,112/9/Product: coat protein VP2 #status predicted <VP2>
 F,112,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 12.1%; Score 482; DB 1; Length 729;
 Best Local Similarity 25.7%; Pred. No. 1.3e-23;
 Matches 197; Conservative 112; Mismatches 314; Indels 144; Gaps 35;

Qy 43 RGLVLPQYKYLGPFGNLDKGEPPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQ 102
 Db 9 RGLTLPQYKYLGPFGNLDKGEPPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQ 68
 Qy 103 RLQEDTSFGNLRGAVPQAKKRVLEPLGLVEGAKTAPGKRPVQSGQEPDSSSGIGK- 161
 Db 69 ETEHAKDYGKIGYFFPA--KRAFRP--KLESTDSPTTSQOPVRRSRKRPKSPKPKR 125
 Qy 162 -----TQOQPAKRLNFGQTDSESVDPDPLGEPPTPAVAGPTTMASSGGAAMA 212
 Db 126 PAPHPIFINLAKKAKGTSNTNSMSSENVQHNPIANGTEL--SATNGESGGGGG--- 181
 Qy 213 DNEGADGVNAGSNM--CDSTWLG--RVITSTRTWALPTYNHLYQOIS--SAST 265
 Db 182 -GGAGAGGVSTGSPFNQTEFOYLGELVRIITAHASRLIHANPHEHTYRIHVLSSES 240
 Qy 266 GAS-----NDNHYFGVSTPMWGFDFNRFRCHFSPPDMQRLINNMNGFRKRLNPLFNIOV 321
 Db 241 GVAQOMVODDHTQMTQVTPMSLIDANAMGVWPNPDMOLISNMTEINLVSEQGIENVVL 300
 Qy 322 KEVT-----TNDGVTTIANNLSTVOVFSDEYQLPYVLSAHQGLCPFPADVFMIPOYGY 378
 Db 301 KTIIESATSPPTKYNNDLTASLWALDNTNLTPTPAAPASSETLGPYMLPTKPTQYRY 360
 Qy 379 L-----TLNNGSQAVGRS-----SFYCLE--YFPSSQMLRTGNF--TFSTYFE 417
 Db 361 YLSCIRNLNPTPTYGOSQOITDSIoTGLHSDIMEFYTENAVJILHLRTGDEFSTGIYHFD 420

Qy 418 EVPFHSVYAHQSGLDR-----LNMPLI--DOVLYINRTONQSGAONKDLFSRGS 467
 Db 421 TKPL--KLTHSMQTNRSGLPPLKLTPTTEGQHPGLTLPANTRKGYHTINNSYTEAT 478
 Qy 468 PAGMSVQPKNMLPQPCYRQQRVSKTKTDNNNSNFTWGASKYINANGRESIINP--GTAMA 525
 Db 479 ----AIRP-----AAGVNTPYNMFYSNGSGPF-----LTPIVPTADT 512
 Qy 526 SHKODEDKFFPMGSMWIFGKESAGASNTALD---NVMITDEEIK--ATNP-----VAT 574
 Db 513 QYNDDEP-----GAIPTMYOYGHGLTTSQELERTFENPQSKCGAPK 557
 Qy 575 ERFETVA--VNFQSSS-----TDPATG--DVHMGALP-----GMWQODRDVYLQGPIMA 620
 Db 558 QQFNOQAPLNENTNNGTLTSPDPIGKSNMFMNTLTNTYGFALTALNNTAVFPNGQIMD 617
 Qy 621 KIPHTD--GHPHSPLMGFGIK--NPPQILIKNTPPANPAEFA--TKRASFITQYST 676
 Db 618 KELDTDKPRHL---VTAPEVCNNPPQGLFVKIAP---NLTDENADSPOQPRITITSN 671
 Qy 677 GQVSEIEEMELQKENSKRMPNPEVOYTSNVAKSANVDFVNDNGLYTEPRPIG 728
 Db 672 FWMKGTLTFTAKMRSNMNMPIOQHTT-----TAENIGNYI--PTNIG 712

RESULT 7

VCPVLP coat protein VP1 - feline panleukopenia virus

N/Contains: coat protein VP2
 C/Species: feline panleukopenia virus, FPLV
 C/Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
 C/Accession: A03701
 R/Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winstan, S.; Hahn, W.
 J. Virol. 55, 574-587, 1985
 A/Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parvovirus
 A/Reference number: A03697; M01D:85265017; PMID:2991581
 A/Accession: A03701
 A/Molecule type: DNA
 A/Residues: 1-727 (CAR)
 A/Cross-references: EMBL:M10824; NID:G333474; PID:AAA47161.1; PID:G333476
 C/Genetics:
 A/Intons: 11/1
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein
 F,114-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 11.1%; Score 444.5; DB 1; Length 727;
 Best Local Similarity 23.7%; Pred. No. 3.7e-21;
 Matches 187; Conservative 117; Mismatches 295; Indels 189; Gaps 35;

Qy 43 RGLVLPQYKYLGPFGNLDKGEPPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQ 102
 Db 10 RGLVLPQYKYLGPFGNLDKGEPPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQ 69
 Qy 103 RLQEDTSFGNLRGAVPQAKKRVLEPLGLVEGAKTAPGKRPVQSGQEPDSSSGIGKT 162
 Db 70 QTDATDMDGKIGHYFFRAKALAPVLTDPDHPISRPK--PTKSKRPPIHFINLAK- 127
 Qy 163 GQPAKRLNFGQTDSESV---DQPIGEPRA--TPAAVPTTMASSGGAAMDNNE 216
 Db 128 -----KKKAGAGVCKEDNQAPMSDGAIVQPDGQPAVNERATSGNGSGGG-----GG 176
 Qy 217 GADGVNAGSNM--CDSTWLGDRVITSTRTWALPTYNHLYQOI-----SS 262
 Db 177 GSGGVGISTGTENNQTEFKFLNGWV--ETANSSRLVHNMPESEYKRVVVVNMNDKTA 234
 Qy 263 ASTGASNDNHYFGVSTPMWGFDFNRFRCHFSPPDMQRLINNMNGFRKRLNPLFNIOYK 322
 Db 235 VKGNMALDTHVQIVTPMSLVDANANGWPNPDMQOLIVNTMSELHIVSTEGQIFNVVLK 294
 Qy 323 EVT-----TNDGVTTIANNLSTVOVFSDEYQLPYVLSAHQGLCPFPADVFMIPOYGY 378
 Db 361 YLSCIRNLNPTPTYGOSQOITDSIoTGLHSDIMEFYTENAVJILHLRTGDEFSTGIYHFD 420

```

Db      295 TVSESATQPEKVVYNNDLTASLWALDLSNNTMFTPAAMSETLGFYPMKPTIPTPMRY 354
Qy      369 ----DVEMIPQGYLTLLNNGSQAAGRS-----SFYCLE-YFPQOMLRTGNF-TFS 413
Db      355 PQMDRTLIPSH-----TGTSPTNTNYHGTDPDVQFYTIENSVPVHLRTGDEFAFGT 408
Qy      414 YTFEEVPHSSVYASQSLDLRLMN--PLIDQYLYLNTQNOGSAQKDLLFSRGSAGM 471
Db      409 FPFDCRP--CRLLHTWQTNALGLRP-----FLNSLPQSEGAITNGDILGVQDQKRGV 459
Qy      472 S-----VQPKN-WLPGPCYRQORVSK-----TKTDNNNSNFTWT 504
Db      460 TQMGNTDYITEATIMRAEYGSAPYFSFASSTQPFKIPIAAGRGGAQTDENQAA---D 516
Qy      505 GASKY---NLNGRESINPGT---AMASHKDEDEKFFPMGCVMI FGKESAGASNTALD- 556
Db      517 GDRYAFGRQHGQKTTTGTGTPERFYIAHODT-----GRYPAGDWIOMINFN 563
Qy      557 NNMITDEEIKATNPVATERFGTVAVNPFQSSSTDPAVDYHMGALPGMWQDRDYVLOG 616
Db      564 NLVTNDNVLLPTDPIC---GKTGINY--TNIFNTYGPLTALNNV-----PYVPNG 610
Qy      617 PIWAKIPHTDGHFHPSPLMGFGFLK-----NPPQILIKNTVPVAN--PPA 660
Db      611 QIMDKFDTD-----LKPRLVNAPFVCCNNCPGOLFVAVAPLITNEYDPDA 657
Qy      661 EFATKFPASFTQYSTQGVSEIEMELQKENSKRWNEVOYTSNYAKSANDVFTVNNGL 720
Db      658 SAMMSR---IVTYSDFPMWKGKLVFKAKLRASHMTNPIQMSIN-----VDNQFN 703
Qy      721 YTEPRPIG 728
Db      704 YL-PNNIG 710

RESULT 8
VCPVME
coat protein VP1 - mink enteritis virus (strain Abashiri)
N:Contains: coat protein VP2
C/Species: mink enteritis virus, MEV
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 12-Apr-1996
A/Accession: B38350
R/Karlatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinag
J. Gen. Virol. 72, 867-875, 1991
A/Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the
A/Reference number: A38350; MUID:9120123; PMID:2016557
A/Accession: B38350
A/Molecule type: DNA
A/Residues: 1-722 <RAR>
A/Cross-references: GB:D00765
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein
F/139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match      11.0%; Score 439.5; DB 1; Length 722;
Best Local Similarity 23.6%; Pred. No. 7,7e-21;
Matches 186; Conservative 117; Mismatches 25; Indels 189; Gaps 35;

Qy      44 GLVLPQGYKYLGPFGNLDKGEVPVNAADAALBHDKAYDQOLKAGDNPYLRYNHADAERQER 103
Db      6 GLVLPQGYKYLGPFGNSLDQGEPTNPSDAAKEHDEAYAYLRSGKNPYLYFSPADQRFIDQ 65
Qy      104 LQEDTSFGGNLGRAVPAQAKRVLLEPLGLVEBAKTAQKRPVQSPQEPDSSSGIGKTG 163
Db      66 TKQATWGGKIGYHFFAKKALAPVLTDTDPNBSPTSPK-PTKRSKPPHIFINLAK-- 122
Qy      164 QQPAKRLNFGQTDSESEVP---DPOPLGEPPA--TPAAVGPPTMAAGGAPMADUNEG 217
Db      123 ----KKKAGAGQVYRBDLALMSDGAQVDPGGQPAVRREKATGSGNSGGGG-----CGG 172
Qy      218 ADQVGNASGWNH-----CDSTWLGDRVYITTSRTYALPTYNHLYKQI-----SSA 263
Db      173 SGGVGISTGTFTNQTGKFLNGVW--EITANSRLVHLNMPSEENYKRVVANNMDKTAV 230

```

```

Qy      264 STGASNDNHYFGYTPMGYEDFNRFCHSPRDMORLNNMGFRPKRLNFKLNIQVKE 323
Db      231 KGNMADDHDTVOQVLTWMSLYDANAMGWBNRPGDQMLIVTMSEGLHVSFEQELFNVLKT 290
Qy      324 VT---TNDGVTTIANNLSTVQYFSDSEYOLPYVLSAHQGL-----PPPPA----- 368
Db      291 VSESATQPTKVVYNNDLTASLWALDLSNNTMFTPAAMSETLGFYPMKPTIPTPMRY 350
Qy      369 --DVEMIPQGYLTLLNNGSQAAGRS-----SFYCLE-YFPQOMLRTGNF-TFSY 414
Db      351 QMDRTLIPSH-----TGTSPTNTNYHGTDPDVQFYTIENSVPVHLRTGDEFAFGT 404
Qy      415 TFEVPHSSVYASQSLDLRLMN--PLIDQYLYLNTQNOGSAQKDLLFSRGSAGM 472
Db      405 FPDCKP--CRLLHTWQTNALGLRP-----FLNSLPQSEGAITNGDILGVQDQKRGV 455
Qy      473 -----VQPKN-WLPGPCYRQORVSK-----TKTDNNNSNFTWTG 505
Db      456 QMGNTDYITEATIMRAEYGSAPYFSFASSTQPFKIPIAAGRGGAQTDENQAA---DG 512
Qy      506 ASKY---NLNGRESINPGT---AMASHKDEDEKFFPMGCVMI FGKESAGASNTALD-N 557
Db      513 DRYAFGRQHGQKTTTGTGTPERFYIAHODT-----GRYPAGDWIOMINFN 559
Qy      558 VMTDEEIKATNPVATERFGTVAVNPFQSSSTDPAVDYHMGALPGMWQDRDYVLOG 617
Db      560 LPTVNDNVLLPTDPIC---GKTGINY--TNIFNTYGPLTALNNV-----PYVPNG 606
Qy      618 IWAKIPHTDGHFHPSPLMGFGFLK-----NPPQILIKNTVPVAN--PPA 661
Db      607 IMDKFDTD-----LKPRLVNAPFVCCNNCPGOLFVAVAPLITNEYDPDA 653
Qy      662 FSATKFPASFTQYSTQGVSEIEMELQKENSKRWNEVOYTSNYAKSANDVFTVNNGL 721
Db      654 ANMSR---IVTYSDFPMWKGKLVFKAKLRASHMTNPIQMSIN-----VDNQFN 699
Qy      722 TEPRPIG 728
Db      700 L-PNNIG 705

RESULT 9
VCPVFP
coat protein VP1 - feline panleukopenia virus (strain 193)
N:Contains: coat protein VP2
C/Species: feline panleukopenia virus, FPLV
C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 12-Apr-1996
A/Accession: B36608
R/Martyn, J.C.; Davidson, B.E.; Studdert, M.J.
J. Gen. Virol. 71, 2747-2753, 1990
A/Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvo
A/Reference number: A36608; MUID:91073139; PMID:2174965
A/Accession: B36608
A/Molecule type: DNA
A/Residues: 1-727 <MAR>
A/Cross-references: GB:X55115
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein
F/144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match      10.8%; Score 431.5; DB 1; Length 727;
Best Local Similarity 23.5%; Pred. No. 2.6e-20;
Matches 185; Conservative 117; Mismatches 297; Indels 189; Gaps 35;

Qy      43 RGLVLPQGYKYLGPFGNLDKGEVPVNAADAALBHDKAYDQOLKAGDNPYLRYNHADAERQER 102
Db      10 RGLVPPGYKYLGRGNSLDQGEPTNPSDAAKEHDEAYAYLRSGKNPYLYFSPADQRFID 69
Qy      103 LQEDTSFGGNLGRAVPAQAKRVLLEPLGLVEBAKTAQKRPVQSPQEPDSSSGIGKT 162
Db      70 QTDARDWGGKIGYHFFAKKALAPVLTDPDPSPTSPK-PTKRSKPPHIFINLAK-- 127

```


A: Molecule type: DNA
A: Residues: 1-143, 'A', 145-718 <AST>
A: Cross-references: EMBL: M12032
C: Superfamily: parvovirus coat protein
C: Keywords: coat protein

Query Match 10.6%; Score 424.5; DB 1; Length 718;
Best Local Similarity 23.7%; Pred. No. 7.3e-20;
Matches 184; Conservative 104; Mismatches 298; Indels 189; Gaps 31;

```
QY 45 LVLPKTYLGPNGLDKGEVNAADAAALHDKAYDQOLKAGNPYLRNHADAERLER 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MVPPKTYLGPNGSLDGEPTNPSDAAKHDEAYDQYIKSGNPYLFSAALQRFIDQ 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 105 QEDTSGNLRGAVFOAKKRVLEPLGLVEBGAKTAPGKKRPVBOGPOBSSSGIGKTG 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 KDAKMGKGVKGYHFFKTKRAPAPGLATDSE--PQ-----TSGVSDAGK 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 -----QOPAKKRLNFG-----QTGDSSEVDDPQPLG-----EPAPTPAAVGPPTM 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 RTRPRAVIFINQARAKKLTSSAAGSSQMSDQTSQPDGNGVHSAARERADGPG-- 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 204 ASGGAPADNNGAGDVGASGNWCHDS--TWLGD--RVITSTRTVALPTYNHLYK 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 GSGGG-----GSGGGGAGVGTSGSYDQTHYRFLGDGWEITLALATRLVLANPKSENYC 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 QI-----SSATGASNDNHFGYSTPMGYPDPNRFHCHSPRDMQRLINNMWGRPRK 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 RIVENHTTDSVKGNAKDDAHQIWPMSLVANAGVWLQPSDWQYICNTMSQLNLVS 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 LNEKFNIOYKEVYTNND---GVTTIANNLSTVOYFSDSEYOLPYLGSAGHGCLEPPF 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 LDEIEFNVLKTYTEBDSGGQALKIYNNDLITACMVAVDNNTLPTPAANSMEITLGFY 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 368 ADVFMIPOGY-----LTLAN-----GSAVGRSSFCLEYFSG--QMLRTG 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 WKFTIASPYRYFVCVDRDLSTVYENEGTIEHNVMGTGPKGMSQFTEITNTOITLLRTG 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 408 NNE--TSTYFEEVPHSSVYASOSLRLNM--PLIDQYLYIARNTQOSASQKOLLFS 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 DEBATGYTFYDTPV--KLTHWTQTNQLDQPLLSTF--PEADTADT--LTA 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 RGSPPAGSVOPKRWL-----PGPCROORVASKTKTDNNNSN 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 QGSRHGAQTQMEVWWSAITRPAQVGFQCPHNDFFASRAGP--FAAPKVADVTQGVDR 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 501 FTWTGASKYNLNGRESIINFGTAMASHKDEDKF-----FPMGVMIFKESAGAS 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 ANGSVYYSYQKQGEWMAAHGAPAPERYTMDETNFGSGRDRDFTQSAPIVPPPLNGIL 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 552 NTALDVMITDEBEIKATNPVATERFGTVAVNFQS--SSTDPAATGVDHAMGALPGMWOD 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 TNM-----NPIGTKN--DIHFSNVNNSYGLPLASH-----P 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 RDVYLOGPIWAK--IPHTDGHFSPSLMGFGILKNPPQILIKNTP--VPANPAFSA 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 SPVYPCQIMDKDLDEHKRLHITAPFV--CKNNAFGMLVRLGNNLTQDQDPNGATL 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 665 TKFASPTQYSTQGVSEIEMWELQKNSKKNMPEVOYTSYAKSANVDFVNNNG 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 SRIVTGYTFPMWKGKLTWRAKLRA-----NTWMNPFYQ-----VSEVDNG 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 12

VCPCVCP
coat protein VP1 - canine parvovirus (strain N)
N: Contains: coat protein VP2
C: Species: canine parvovirus, CPV
C: Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 12-Apr-1996
C: Accession: B29962
R: Reed, A.P.; Jones, E.V.; Miller, T.J.
J. Virol. 62, 266-276, 1988
A: title: Nucleotide sequence and genome organization of canine parvovirus.

A: Reference number: A29962; MUID: 88062992; PMID: 2824850
A: Accession: B29962
A: Molecule type: DNA
A: Residues: 1-748 <RE>
A: Cross-references: EMBL: M19296
C: Gene(s):
A: Introns: 26/3
C: Superfamily: parvovirus coat protein
C: Keywords: coat protein
P: 165-748/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.6%; Score 421.5; DB 1; Length 748;
Best Local Similarity 23.0%; Pred. No. 1.2e-19;
Matches 184; Conservative 114; Mismatches 287; Indels 215; Gaps 34;

```
QY 44 GLVLPKTYLGPNGLDKGEVNAADAAALHDKAYDQOLKAGNPYLRNHADAERLER 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 GLVPPKTYLGPNSLDDGEPTNPSDAAKHDEAYDQYIKSGNPYLFSPADQRFIDQ 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 104 LOEDTSFGNLRGAVFOAKKRVLEPLGLVEBGAKTAPGKKRPVBOGPOBSSSGIGKTG 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 TKDAKMGKGVKGYHFFKTKRAPAPGLATDSE--PQ-----TSGVSDAGK 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 OOPAKKRLNFGQTGDSSEVP---DQPLGEPPA--TPAAVPTTASGGA PMADNNG 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 ---KKKAGAGYKRDNLAPMSDGA VQPDGQDPAVNERATSGNGSGGG-----GGG 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 218 ADGCVNASGNW-----CDSTWLADRVITSTRTVALPTYNHLYKQI-----SSA 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 SGGVGI STGTFTNNQTEFEKLENGW--EITANSRLVHLMSESENYRVRVANNMDKTAV 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 STGASNDNHFGYSTPMGYPDPNRFHCHSPRDMQRLINNMWGRPRKRLNFKLNIQVE 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 NGNMA LDDIHAQVTPMSLV DANAGVWPNFGDQULIYVMSGLHVS EOEIEFNVLK 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 324 VT---TNDGVTTIANNLSTVOYFSDSEYOLPYLGSAGHGCLEPPF-----P 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 VSESATQPTKYNNNDLITACMVAVDNNTLPTPAANSMEITLGFYPMKPTIPTMRYF 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 369 ---DVFMIPOGYLTLNNGSQA VGRS-----SFYCLE-YFSPQMLRTGNF-TSEY 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 QMDRTLIPSH-----TGTSGPTNTYHGTDPDDVQFYTIEHSVPHLLRTGDEPATGTF 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 TFEFEEVPHSSVYASOSLRLNM--PLIDQYLYIARNTQOSASQKOLLFSRGPAGMS 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 PFOCKP--CRLTHWTQTNALGLP-----FLASLPQSEGA TNFGDI-----GY- 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 473 VOPKRWLPGPCYRQGVSKTKTDNNNSN-----TWGASKYNLNGRESIINP-GT 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 -----QODKRGVITQMGNTNVTIETATIRPAVGVSAVYFSFASSTQGPFT 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 523 ANASHK-----DDEDKFPFMSGVMIFKESAGASNALD----- 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 PIAAGRGAQTYENQALDDPRY-----AFRGQHGQKTTTGTGPERFTYIAHODTR 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 557 -----NVMITDEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMGALPG 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 YPEGMTIQTINENLPTNDNLPTDPIG--GKTGINY--TNTFNTYGLPLALNNVP- 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 MWQDRVYLOGPIWAKIPHTDGHFSPSLMGFGILKNPPQILIKNTP--VPANPAFSA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 -----PVPNGQIMWKEFDTD-----LKRPLHVNAPVQCQNGCGQIFVAV 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 652 TPVPAN--PPAFSA TKFASPTQYSTQGVSEIEMWELQKNSKKNMPEVOYTSYAKS 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 APVLTTEYDPPDASANMSR---IVTYS DFWMKGLVFKAKLRASH TWNIDIQMSIN--- 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 ANVDFTVDNNGLYTEBPRIG 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 -----VDNQFNIV-PSNIG 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 13

VCPVCD
coat protein VP1 - canine parvovirus (strain CPV-d)
N/Contains: coat protein VP2
C/Species: canine parvovirus, CPV
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C/Accession: A31163
C/Paritoh, C.R.; Aquadro, C.F.; Carmichael, L.E.
Virology 166, 293-307, 1988
A/Title: Canine host range and a specific epitope map along with variant sequences in th
A/Reference number: A31163; MUID:85020796; PMID:3176341
A/Accession: A31163
A/Molecule type: DNA
A/Residues: 1-737 <PAR>
A/Cross-references: EMBL:M23255; NID:G333467; PIDN:AAA47158.1; PID:G333468
C/Genetics:
A/Insertions: 26/3
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein
F/584-737/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.2%; Score 407.5; DB 1; Length 737;
Best Local Similarity 22.8%; Pred. No. 9.8e-19;
Matches 181; Conservative 112; Mismatches 301; Indels 201; Gaps 32;
42 GRGLVLPQYKYLGPFGNLDKGEPPVAAADAALEHDKAYDQOLKAGDNPYLRYNHADAEPQ 101
19 GEGKDLITYKYLGGNSLDGEPTNPDAAKEHDEAAVLRSGKPYLYFSPADQRFI 78
102 ERLQEDTSPGNLGRAVFOAKKRVLEPLGLVEBQAKTAPGKRPVEGSPQEPDSSSGIGK 161
79 DQTDADKMGKIGHYEFRAKKAIAVLTDPDHPSTSRPTK-PTKSKRPPIIFIMLAK 137
162 TGGQPAKKRLNFGQTDGSESV-DEPRLGEBPA--TPAVGPPTMASGCGAPMADNN 215
138 -KKKAGAGVYKRDNLAPMSDGAIVPDGCGPAPVNERATSGSGSGG- -G 185
216 EGADGVNAGSNM- -CDSTWLGDRIYITSTRWALFTYNNHLYKOI- -S 261
186 GGSAGVIGSTGTFNNQNEFKLENGW- -ETANSSRLVHLMPSESEYRVRVNNMDKT 243
262 SASGASNDNHYFGYSTPWGYFDNRFHCHFSPPDMQRLINNMGFRPKLNFLENIQV 321
244 AVNQMALDIDIAOIVTPMSLVLDANAMGVNPPGMQILVNTMSEHLVSPFEOEIFNVVL 303
322 KEVT- -TNDGVTTIANLSTVOVFSDEYQLPYVLSAHOGL- -PPPPA- - 368
304 KTVSESATQPTKYYNNDLTASLVAALDSNNTPTPAAMSESLGTFPKPPTIPTPKRY 363
369 - -DVFMIPOYGYLTLNNGSAVGRS- -SFYCL- -YFPGQMLRTGNNP- -TF 412
364 YFQMDRTLIPSH- -TGTSGRPTNIYHGTDPPDVQFYTIENSVPVHLRTGDEPATG 417
413 SYTEBEVPFSSSYAHSGSLRLM- -PLIDQYLYLARTQNGSSAQKDLIFRSGSAG 470
418 TFPDCKR- -CRLLTHWQTNRALGLP- -FLMSLPQSBGATNFGDI- -G 460
471 MSVPKMWLPQPCYRQGVSKTKTDNNNSNF- -TWGASKNLNGRESIINP- 520
461 V- - - - -QDDKRGVTOGNTNITTEATIMRPAVGVSAPIYFSEASTGCPF 506
521 - - - - -GTAMASHKDEDEKFPMSGVMI FGKESAGASNTALD- - - - - 556
507 KTIPLAAGGAGQTDENQADGNPRYAFGRHGQKTTTGTGERFTYLAHQDGRYPEGD 566
557 - - - - -NMVITDEBEIKATNPVATERGTAVANFOSSSTDPAGDVAMALGGMWQD 609
567 WIONINFPLPTVNTNVLPTDPIG- -GKTGINV- -TNIPNTYGPLALANNVP- - - - - 614
610 RDVLYOGPIAKTIDHTDGHFHPSPLMGFGK- - - - -NPPOLLINKTPVPA 656
615 - - - - -PVPNGQIMKEPDTD- - - - -LKPRLHVNAPVCCQNNCGDLFVYAPRLT 660
657 N- - - - -PPAEFSATKBPASPIYTGQGVSEIEMELOKENSKRMNDEVQYTSNYSKANDVF 713

Db 661 NEYDPPASAMSR- - - - -ITYSDPFMWKGLVFAKALRASHITWNPIDQMSIN- - - - - 707
QY 714 TVDNNGLYTERPPIG 728
Db 708 -VDNQFNYV-PSNIG 720

RESULT 14

VCPV2
coat protein VP1 - parvovirus H1
C/Species: parvovirus H1
A/Note: host Homo sapiens (man)
C/Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
C/Accession: A03699
R/Rhodes III, S.L.; Paradiso, P.R.
J. Virol. 45, 173-184, 1983
A/Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid
A/Reference number: A03695; MUID:83112183; PMID:6823009
A/Accession: A03699
A/Molecule type: DNA
A/Residues: 1-722 <RHO>
A/Cross-references: EMBL:X01457; EMBL:J02198
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein

Query Match 10.1%; Score 404; DB 1; Length 722;
Best Local Similarity 23.3%; Pred. No. 1.6e-18;
Matches 175; Conservative 108; Mismatches 290; Indels 178; Gaps 30;
57 NGDLKGEPPVAAADAALEHDKAYDQOLKAGDNPYLRYNHADAEPQRLQEDTSFGNLR 116
12 NSLDGEPTNPDAAKEHDEAVDQYIKSGKNPYLYFSPADQRFIDQTDKADKMGKVG 71
117 AVQAKKRVLEPLGLVEBQAKTAPGKRPVEGSPQEPDSSSGIGKQCO- -PA- - - - - 167
72 YFRTKAFAFPKISTDSE- - - - -FG- - - - -TSVSRPKRTKPPAHIFVQ 112
168 - - - - -KKRLNFGQTDGSESVDPQPLGEPATPAVGPPTMASGCGAPMADNNEGADGVN 223
113 ARAKKRAASLAQORTITMSDGTETNOPTGINAKAVERSADGGG- -SGGSGSGGGIGV 171
224 ASGNMHCDSW- -LGDRIYITSTRWAL- - - - -PTYNNHLYKOISSASTGA 267
172 STGYNDQTYKFLGQGWAITAHASRLHLGMPSENYCRYVTHNNQTLGHSTKYKGM 231
268 SNDNHYFGYSTPWGYFDNRFHCHFSPPDMQRLINNMGFRPKLNFLENIQVKEVT- 326
232 AYDTHQIW- -TPMSLVLDANAMGVNPPGMQILVNTMSEHLVSPFEOEIFNVVL 303
327 - - - - -NDGVTTIANLSTVOVFSDEYQLPYVLSAHOGL- - - - -PPPPA- - - - -VFM 373
291 QGAGQALIKYNNNDLTACMNVALDSNNTPTPAAMSESLGTFPKPPTIPTPKRY 363
374 PÖGYLYTLNNGS- - - - -AVG- - - - -RSSFYCL- -YFPGQMLRTGNNP- - - - -SYTEBEVP 420
351 PROLSVTSNAGTQITDITIGEPQALNSQFTEIENITLPTLLRTGDEFTGYIFNTDP 410
421 FHSYA- - - - -HQSLRLM- -PLIDQYLYLARTQNGSSAQKDLF- - - - - 463
411 LKLTHTWQTNRLHACLOGITDLPSTPATASLPLANGREFSTQONNVYTEALRTTAPQ 470
464 - - - - -SRGSPAGMSVOPKMWLPQPCYRQGVSKTKTDNNNSNFTWTGASKYKL 511
471 IGMQPHDNPEARNGEPFKVPVP- - - - -LDITABEDHAN- - - - -GALRFN 513
512 NGR- - - - -ESINPSTA- - - - -MASHKDEDEKFPMSGVMI FGKESAGASNTALDVM 559
514 GKQHGEMAKQGAPEBYTDAIDSAAGRDTACFV- - - - -QSAPISIPPNQOI 563
560 ITDEBEIKA- - - - -TNVVALERFGTAVANFOSSSTDPAGDVAMALPGMWDQDVYLOGPI 618
564 LQREDAIAGRTNNHYTNVFNVSYPPLSAFPHDP- - - - -IYPNGOI 603

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 14.5704 Seconds
(Without alignments)
2095.115 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989

Sequence: 1 MADGYLPLWLEDNLESGIR.....NNGLYTEPRPIGRVLRPL 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	54.6	504	1	COA3_AAV2
2	798.5	20.0	673	1	COAT_PAVBO
3	600.5	15.1	781	1	COAT_PAVHB
4	482	12.1	729	1	COAT_PAVP9
5	482	12.1	729	1	COAT_PAVPN
6	480	12.0	729	1	COAT_PAVK
7	444.5	11.1	727	1	COAT_PAV
8	441.5	11.1	749	1	COAT_PAVN
9	437.5	11.0	722	1	COAT_MEVA
10	431.5	10.8	727	1	COAT_PAV19
11	430.5	10.8	727	1	COAT_PAVCB
12	424.5	10.6	718	1	COAT_MOMIM
13	421.5	10.6	748	1	COAT_PAVCD
14	407.5	10.2	737	1	COAT_PAVCD
15	404	10.1	722	1	COAT_PAVHH
16	386.5	9.7	722	1	COAT_PAVTC
17	377	9.5	716	1	COAT_MOMIV
18	234	5.9	587	1	COAT_PAVT3
19	207.5	5.2	809	1	COAT_DSDNV
20	201	5.0	811	1	COAT_GMDNV
21	199.5	5.0	810	1	COAT_JCNDV
22	195	4.9	584	1	COAT_PAVC2
23	186.5	4.7	647	1	COAT_PAVG
24	143.5	3.6	648	1	WHN_MOUSE
25	131.5	3.3	1446	1	IE18_PAVKA
26	126.5	3.2	1461	1	IE18_PAVIF
27	124.5	3.1	2493	1	CVAA_USTMA
28	123	3.1	648	1	WHN_HUMAN
29	123	3.1	1142	1	ENAM_PIG
30	121	3.0	880	1	SYV_EACST
31	121	3.0	1113	1	N116_YEAST
32	120	3.0	1379	1	YFF9_SCHPO
33	119.5	3.0	1394	1	HAP_HAERIN

34	119.5	3.0	2004	1	YP73_YERPE	08z412 yersinia pe
35	119	3.0	667	1	ITP2_HUMAN	P15884 homo sapien
36	117.5	2.9	1076	1	Nup1_YEAST	P20676 saccharomyc
37	116.5	2.9	642	1	ITP2_CANFA	P15881 canis famil
38	114.5	2.9	802	1	PAC_BACME	060136 bacillus me
39	113.5	2.8	1324	1	IR52_HUMAN	09y4h2 homo sapien
40	113	2.8	837	1	HFC2_HAEIN	P45997 haemophilus
41	113	2.8	1484	1	CES2_HUMAN	09bxt3 homo sapien
42	112	2.8	774	1	PECA_ECOLI	P13036 escherichia
43	111.5	2.8	2944	1	CA17_HUMAN	002388 homo sapien
44	111	2.8	454	1	NOE2_HUMAN	095897 homo sapien
45	111	2.8	635	1	HMLA_DROME	P10105 drosophila

ALIGNMENTS

RESULT 1	COA3_AAV2	STANDARD;	PRT;	504 AA.
ID	COA3_AAV2			
AC	P03135;			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-NOV-1991 (Rel. 20, Last annotation update)			
DE	Probable coat protein 3.			
OS	Adeno-associated virus 2 (AAV2).			
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
OX	NCBI_TaxID=10804;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=83164299; PubMed=6300419;			
RA	Srivastava A., Luby B.W., Berns K.I.;			
RT	"Nucleotide sequence and organization of the adeno-associated virus 2 genome."			
RL	J. Virol. 45:555-564 (1983).			
CC	CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	CC			
DR	EMBL; J01901; AAA2376.1; -.			
DR	PIR; A03698; VCPV3A.			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
KW	Coat protein.			
SQ	SEQUENCE 504 AA; 56366 MW; 758999B017052B6B2 CRC64;			
Query Match	54.6%; Score 2177.5; DB 1; Length 504;			
Best Local Similarity	82.5%; Pred. No. 2.5e-128;			
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;				
QY	203 MASGGAPADANNNEGADVGNGASGNMHCSTWLGDRVITTSRTTALPTNNHLYQOIS 262			
DB	1 MATGSSGAPADANNNEGADVGNGSSGNMHCSTWLGDRVITTSRTTALPTNNHLYQOIS 60			
QY	263 ASTGASNDNHYFSGYSTPMGTFDFNRFCHFSPPDMORLINNMMGFAPKRLNFKLFNIQVK 322			
DB	61 QS-GASNDNHYFSGYSTPMGTFDFNRFCHFSPPDMORLINNMMGFAPKRLNFKLFNIQVK 119			
QY	323 EYTTNGVTTIANNLSTVQVPSDSYQLPYVIGSAHOGCLPFPADVMTPOYGYTLN 382			
DB	120 EYVQNGVTTIANNLSTVQVPSDSYQLPYVIGSAHOGCLPFPADVMTPOYGYTLN 179			
QY	383 NSGQAVGRSSFYLEYFSPOMLRTGNMFESYFEEVPPHSSVAHSOSLDRLNNPLIDQY 442			
DB	180 NSGQAVGRSSFYLEYFSPOMLRTGNMFESYFEEVPPHSSVAHSOSLDRLNNPLIDQY 239			
QY	443 LYVINTQNGSGAQNKDLLFSRGSFAGMSVOPKNNLPGPCYRQGRVSKTKTDNNNSNFT 502			

```

Db 240 LYLSTRTNTPSGTTSQSLRQFSQAGASDIRDQSRNWLPGPCRYQRYVSKTSADNNNSEYS 299
Qy 503 WTGASKYNLRRESIIINPGTAMASHKODEDEKFFPMGSGMITGKSAGASNTALNNWITD 562
Db 300 WTGATKTHLNKRDLSLVN-AMASHKODEEKFPPQSGVLPGKQSEKTNVNIKMTD 357
Qy 563 EEEIKATNPVATFEFGTAAVAVFQSSSTDPATGDVHAMGALPGMWODRDVYLQGPIMAKI 622
Db 358 EEEIGTNPVATEQYGSVSTLQKGNRQAAATADVNTQGVLPFGMWODRDVYLQGPIMAKI 417
Qy 623 PHTDGHFHPSPPLMGFGGLKNDPPOLLKNTVPANPPAFSAYTASFTTQSTG 677
Db 418 PHTDGHFHPSPPLMGFGGLKNDPPOLLKNTVPANPPAFSAYTASFTTQSTG 472

RESULT 2
COAT_PAVBO STANDARD; PRT; 673 AA.
ID COAT_PAVBO
AC P07297; O84374;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP2 [Contains: Coat protein VP3].
OS Bovine parvovirus (BPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=10784;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87061184; PubMed=3783814;
RA Chen R.C.; Shull B.C.; Moses E.A.; Lederman M.; Stout E.R.;
RA Bates R.C.;
RT "Complete nucleotide sequence and genome organization of bovine parvovirus."
RL J. Virol. 60:1085-1097(1986).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14363; AAB59848.1; -
CC DR EMBL; M14363; AAB59849.1; -
CC DR PIR; A26104; VCPVBS.
CC DR HSSP; P30129; ADPV.
CC DR InterPro; IPR001403; Parvo_coat.
CC DR Pfam; PF00740; Parvo_coat; 1.
CC KM Coat protein; Glycoprotein.
CC FT CHAIN 1 673 COAT PROTEIN VP2.
CC FT 138 673 COAT PROTEIN VP3.
CC FT CARBOHYD 343 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 344 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 346 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT DOMAIN 163 187 GLY-RICH.
CC SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 20.0%; Score 798.5; DB 1; Length 673;
Best Local Similarity 28.6%; Pred. No. 2.1e-42;
Matches 214; Conservative 105; Mismatches 275; Indels 155; Gaps 21;

```

```

Db 112 LYPFASNGKAKKANREPAPSTNQOMWEVSNIDIPNDAQNPTELATRSVVGSGVGGG 171
Qy 208 GAPMADNNEGADVGNAGNMGHCDSTWLGDRVITTTSTRTALPTNNHLYKQISSASTGA 267
Db 172 -----RGSSVGVSTGTGTTGTFSENI VVTKNTROFICDKGHLYKS-EVANTGD 222
Qy 268 SUNDHFGVSTPGVDFENRFFCHFSRPRDQRLINNMGFRPRKLNPKLFNIOKEVTIN 337
Db 223 TARKQY-AITTPSYNRFNOYSSHFSPNDQHLVNDYERRPRPAMLVRYVNLQIKQIMTD 281
Qy 328 DGVTTIANN-LTSTVOVFSDEYQLPYLGSAHQGLPPPADVEMIPOYGYLTL----- 381
Db 282 GAMGTVYNNDLTAGMHI FCDGDRYPYQHPMDQCMPELPSIWEHPYAYIPADISVY 341
Qy 382 -NNGSQAVGRS-----SFYCLEFPQOMLRTGNPFSTFYFEVFPFHSSVAHSQSIDRLM 435
Db 342 DNNVTNVEEHLKGVPLVYLENSDHEVLNNG-----RTY 376
Qy 436 NPLIDQVLYYLNRTON-----QSGAKNDLFFSRGSP-----AGMSYQPKNMLP 480
Db 377 RIYQLRLMNDKQHHIQASDVOSTGQKQNLIQKQPNKORFOAAALRTSNMS 436
Qy 481 GPCTYQOQVSKITDNNNSNFTWTGASKYNLRRESIIINPGTAMASHKODEDEKFFPMGV 540
Db 437 GP-----GIARGTHNATLQOSAGALVTWVINGAD-----VSGV 470
Qy 541 M-----FGKE-----SAGASNTALDNNMITDEEIKATNPATE 575
Db 471 RAVRVGSTDPIYIGGQOPESSDLRLRYASAAAGQNPILN-----AARH 516
Qy 576 RFGTAAVAVFQSSSTDPATGDVHAMGALPGMWODRDVYLQGPIMAKI PHTDGHFHPSPPLM 635
Db 517 TFRREARKLITSGNADGQYKEMWMLPNQMWASAPISRNPIWVAVPRNRKTLTDTD 576
Qy 636 GGFGLKNDPPOLLK-NTVPANPPAFSAYTASFTTQSTGQVSVIEIEMELQENSK 633
Db 577 GSIPMSHPPTGIFIKARIPVQNGD-----SFLNLYYGVGVSCVEMVEVERKGT 627
Qy 694 RANPEVQYTSNVAKSANVD-FTVDNNGLY 721
Db 628 NMRPEYVHS---ATNMSVDVYTTNNAGVY 653

RESULT 3
COAT_PAVHB STANDARD; PRT; 781 AA.
ID COAT_PAVHB
AC P07297;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Probable coat protein Vp1.
OS Human parvovirus B19.
OC Viruses; ssDNA viruses; Parvoviridae; Erythrovirus.
NCBI_TaxID=10798;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=Isolate AU.
RA MEDLINE=86200451; PubMed=3701931;
RA Shinde R.O.; Blundell M.C.; Cotmore S.F.; Tattersall P.; Astell C.R.;
RT "Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis."
RL J. Virol. 58:921-936(1986).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13178; AAA68687.1; -

```

DR	PIR: A24299; VCPYL9.
DR	InterPro: IPR001403; Parvo_coat.
KW	Plam; Pf00740; Parvo_coat; 1.
FT	Coat protein; Glycoprotein.
FT	CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE	781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;
Query Match	15.1%; Score 600.5; DB 1; Length 781;
Best Local Similarity	26.6%; Pred. No. 5.3e-30;
Matches 182; Conservative	95; Mismatches 300; Indels 107; Gaps 20
QY	47 LPGYKYLGFENGLDKGEVNAADAALBHDKAYDQOLKAGDNPYLRYNHDAEFOERLOE 106
DB	125 LPGNNYVPGNGLQGPQSAVDASARLHDFRYSQALKIGLIPYTHWTVADELLKNIKIN 184
QY	107 DTSGGNLGRAVFOAKRRLVELGLVEGAKTAPKPKRPVYESPOEPDSSSIGTGQOP 166
DB	185 ETGFQAOVVKYF-----TLKGAAPVAHFQ-----GSLPEVP 217
QY	167 AKKRLNFGQTDGDESEVPDPPLGEPBPATPAAVGPPTMASGGGAPMADNEGADGVNMSG 226
DB	218 AYNA-----SEKYPMTSVNSAEASGA-----GGG-----GNSVYS 250
QY	227 NMHCDSTWLADRVITTSRTYALPTYNNHLKYKISSASTGASND-----NHRYGY 276
DB	251 MMSGATYSAHSVCTFSRQFLIPDPEHHYKVFSPAASSCHNAGSKEAKVCTISPIMGY 310
QY	277 STPMGYFPNRFHCFSPRDMQRLINNMGFFPKLNFLENIQYKEVT--TNDGYTTIA 334
DB	311 STPMRYLDFNALNLFPSLEPFLHLENGSIALDALTYVISIAVKDYDKTGGGV-QVT 369
QY	335 NNLTSTVOVFSEKQLPYVLGSAHQGLCPPEPAVPMIROYGYTLNN-GSQAVG---- 389
DB	370 DSTTGRLCLVDHEIKYRYVVLGGQODTLAPELPIWTPPYPAIYLVGDVNTQGISGDSK 429
QY	390 -----RSSFYCLEEYPSQMLRTGNNFTFSYTEEYVFFHSVYASOSLDRLANNPLIDQYLY 444
DB	430 KLASBESAFYVLEHSSFQLLGCGTAHMSYKRPVPPEMLBEGCHFYEMVNPV---YGS 486
QY	445 YLANFTQNSGSAQNKOLLFSKGSPLKMSVQPKMLPGPCYRQQRYSKTYTDNNNSNFTWT 504
DB	487 RLGVADDTLGGCBPKFRSL-----THRDHAIQPCNFMFPGPLVNSVSTEGDSSNTGAKALT 541
QY	505 GASKYNLNGRSLINRG--TAMASHODEDEKFPFMGSMIFGESASGASTALDNY----- 558
DB	542 GLSTGTSQNTIRSLRPGVSPQPTHMDTDKYVTGINALSHGQTTG--NARDKEYQGV 598
QY	559 --MITDEEIKATNVATERFGTVAVNFOSSTSDATGCVHMGALPGKVMODRDVYLOG 616
DB	599 GRFPNEKQKQLQGLQGLNMHTY-----FPNKTOQYTDQIE-RPLMVGSVNRRRLAHYS 651
QY	617 PIMAKIPIPTDGHFHS-PLMGFGGLKNPPQILINTPYANPAPAEFSATKFASTFYOYS 675
DB	652 QLMSTKIPMLDSSFKTQFALGCGMGHQPPEPQFLFK--ILPQSGPIGIGIKSMGITTLLVQYA 709
QY	676 TGVQSVLEIEMEL-QKENSKRNNPE 698
DB	710 VGIMTVMTFKLGPRKATGRRNPQ 733
RESULT 4	
COAL_PAVP9	
ID_COAL_PAVP9	STANDARD; PRT; 729 AA.
AC	P33484;
DT	01-FEB-1994 (Rel. 28, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Coat protein VP1 [Contains: Coat protein VP2].
OS	Porcine parvovirus (strain 90HS) (ppv).
OC	Virusae; ssDNA virusae; Parvoviridae; Parvovirinae; Parvovirus.

Query Match	Beet Local Similarity	Matches	Score	DB 1	Length	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416
-------------	-----------------------	---------	-------	------	--------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

QY 580 VA-VNFOSS-----TDPATG--DVHMGALP-----GWWQDRDYVLOGPIWAKIPIHT 625
 DB 563 QAPLNLNTNNGTLLPSDPISGKPMHFMNTLNTYGPLTALNNTAPVPNGQIWDKELDT 622
 QY 626 D-GHFPSPMLGFGGLK-NPPQILIKNTVPANPPAFESA-TKFAFIYOSTGQSV 681
 DB 623 DLKRLH---VTAFFVCKNNPPGQLFVYLAIP---NLTDFFNADSPQPRITTSNFMWKG 676
 QY 682 EIEWELCKENSKRNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIG 728
 DB 677 TLFTAKRSSNMNPIQOHTT-----TAINIGNYI-PTNIG 712

RESULT 5
 COAL_PAVPN STANDARD; PRT; 729 AA.
 AC P18546; Q89816;
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [contains: Coat protein VP2] (Version 1).
 OS Porcine parvovirus (strain NADL-2) (PPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=10797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021005; PubMed=2219713;
 RA Vasudacharya J., Basak S., Srinivas R.V., Compans R.W.;
 RT "The complete nucleotide sequence of an infectious clone of porcine parvovirus, strain NADL-2.";
 RT Virology 178:611-616 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90085785; PubMed=2596019;
 RA Vasudacharya J., Basak S., Srinivas R.V., Compans R.W.;
 RT "Nucleotide sequence analysis of the capsid genes and the right-hand terminal palindromic of porcine parvovirus, strain NADL-2.";
 RT Virology 173:368-377 (1989).
 CC -1: SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1: MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1: SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements) or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; M38367; AAA46919.1; -;
 DR EMBL; M38367; AAA46921.1; -;
 DR EMBL; M32787; AAA46917.1; -;
 DR EMBL; M32787; AAA46918.1; -;
 DR PIR; B33743; VCPVNA.
 DR HSSP; P30129; 4DPV.
 DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein; Glycoprotein.
 FT CHAIN 1 729 COAT PROTEIN VP1.
 FT CARBOHYD 151 729 COAT PROTEIN VP2.
 FT CARBOHYD 172 729 COAT PROTEIN VP2.
 FT CARBOHYD 198 729 COAT PROTEIN VP2.
 FT CARBOHYD 282 729 COAT PROTEIN VP2.
 FT CARBOHYD 330 729 COAT PROTEIN VP2.
 FT CARBOHYD 433 729 COAT PROTEIN VP2.
 FT CARBOHYD 471 729 COAT PROTEIN VP2.
 FT CARBOHYD 573 729 COAT PROTEIN VP2.
 FT CARBOHYD 604 729 COAT PROTEIN VP2.
 FT CARBOHYD 651 729 COAT PROTEIN VP2.

FT DOMAIN 175 190 GLY-RICH.
 SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80DB42 CRC64;
 Query Match 12.1%; Score 482; DB 1; Length 729;
 Best Local Similarity 25.5%; Pred. No. 1.1e-22;
 Matches 197; Conservative 111; Mismatches 310; Indels 154; Gaps 35;

QY 43 RGLVLEGYKILGFPNGIDKGEPPVNAADAAALHEHDKYDQOLKAGDNPYLEYNADAEFOE 102
 DB 9 RGLTLEPYKTLGKNSLIDQSEPTNPSDAKKEHDEAYDKIKSGKDPYFFSADEKFIK 68
 QY 103 RLQEDTSFGGNLGRAVFOAKKRVLEPLGLVEBAGTAPAKKRPVEQSPQBDSSGIGK- 161
 DB 69 ETEHADYGGKIHYPFRARAPAKL---SETDPTTSQGPFRSFRGHSKPGKR 125
 QY 162 -----TGQPAKKRLNFGOTGSESPDPOPLGEPPATPAAGPTTMAAGGAPMA 212
 DB 126 PARRHIFINLAKKRAKGTSTNNSMSSEVQHNPIAGTEL-SATGNEGGGGGGG--- 181
 QY 213 DNNEGADGVGNASG--NMHCDSTWLDG--RVITTSRTWALPTYNHLYKOIS--SAST 265
 DB 182 -GGRGAGGVSVSTGTFNNQTEFQYLGGLVRIYAHASRLIHLNMPHEHYKRIHVLSSES 240
 QY 266 GAS-----NDNHYEGYSTPWGYPDFNRFCHFSRDMQRLINNNGRPKRLNFKLFINIQ 321
 DB 241 GVAQWQDDAHQWTPMGLIDANAGWFWNPADWQLISNNMTEINLVSEGEIENVL 300
 QY 322 KEVT---TNDGVTIANLITSTQVSDSEYQLPYVLGASHQCLPPFPADVFMIPQY 378
 DB 301 KITTESATSPPTKIYNNDLASLMLVALLDTNNLTPTPPAARSRSTLGPPLPKPTQYR 360
 QY 379 L-----TLNNGQAVGRS-----SFYCLE-YFPQOMLRGTGNF-TFSYTFE 417
 DB 361 YLSICRNLNPTTYTGSGQQTDSIQGLHSDIMFYITENAVPHLRLTGDEFSTGIYHD 420
 QY 418 EVFPHSSYAHSGSLDR-----LNNPLI--DQYLYYINRTQNGSSAQNKDLFSRGS 467
 DB 421 TKPL--KLTSWQTNLSGLRPLKLTPEPTGEGHQHGTLPANTRKGYHOTINNSYEAT 478
 QY 468 PAKMSVQPKWLGPCPCROQRVSKITDNNNSFTMTGASKYVNLNREGSLNF--GTAMA 525
 DB 479 ---AIRP-----AQVGNTPYNNFEYSNGPF-----LTPIVPLADT 512
 QY 526 SHKDEDKFPFPMGWMIFGESAGASNTALD---NMMITDEEIK--ATNP-----VAT 574
 DB 513 QYNDDEPN-----GALRFMDYGHGHLTSSQGLERTFPPQKCGRAPK 557
 QY 575 ERFETVA-VNFOSS-----TDPATG--DVHMGALP-----GWWQDRDYVLOGPIWA 620
 DB 558 QQFNOQAPLNLNTNNGTLLPSDPISGKSNMFMNTLNTYGPLTALNNTAPVPNGQIWD 617
 QY 621 KIHTTD-GHFPSPMLGFGGLK-NPPQILIKNTVPANPPAFESA-TKFAFIYOST 676
 DB 618 KELDTDLKRLH---VTAFFVCKNNPPGQLFVYLAIP---NLTDFFNADSPQPRITTSN 671
 QY 677 GQVSEIEWELCKENSKRNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIG 728
 DB 672 FPMKGLITFTAKRSSNMNPIQOHTT-----TAINIGNYI-PTNIG 712

RESULT 6
 COAL_PAVPK STANDARD; PRT; 729 AA.
 AC P52501;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [contains: Coat protein VP2].
 OS Porcine parvovirus (strain Kresze) (PPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=73487;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=96183900; PubMed=8642680;
 RA Bergeron J., Hebert B., Tjissen P.;
 RT "Genome organization of the Kresze strain of porcine parvovirus:
 RT identification of the allotropic determinant and comparison with
 RT those of MAD-2 and field isolates.";
 RL J. Virol. 70:2508-2515(1996).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U4978; AAC40230.1; -;
 DR EMBL: U4978; AAC40231.1; -;
 DR HSSP: P30129; ADPV.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 KM Coat protein: Glycoprotein.
 FT CHAIN 1 729
 FT CHAIN 1 729
 FT CHAIN 1 729
 FT CARBOHYD 172 729 COAT PROTEIN VP2.
 FT CARBOHYD 172 729 COAT PROTEIN VP2.
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 729 AA; 80835 MW; B6345BFA0568A16 CRC64;

Query Match 12.0%; Score 480; DB 1; Length 729;
 Best Local Similarity 25.6%; Pred. No. 1.5e-22;
 Matches 196; Conservative 112; Mismatches 315; Indels 144; Gaps 34;

QY 43 RGLVLPKYLKPGNGDKGEPVNAADAALEHDKVDOOLKADNRYLYNNADEFOE 102
 DB 9 RGLVLPKYLKPGNSLDGSEPTNPDAAKEDHEDIDKIKSKNYPFFSADDEFIK 68
 QY 103 RLQEDTSFGNLRGAVFOAKRVLPLGLVEBGAKTAPGKRPVEGSPQEDSSSGIGK- 161
 DB 69 ETEHAKDYGGKIGHYFPAKRAFAFKL---SETDSPTTSQGPVRRSRPHGSKPPGKR 125
 QY 162 -----TGQOPAKRLNFGOTGSESPDDPOLCEPAPTPAANGPTTMAAGGAPMA 212
 DB 126 PABRHIFINLAKKAKKAKTSTNSNSMSENOHNPINAGTEL-SATGNESGGGGGGG- 181
 QY 213 DNNEGADGVGNASGNWH--CDSTWLGD---RVITTSRTMALPTYNHNLKQIS--SAST 265
 DB 182 -GGRGAGGVVSGSFNNQTEFOYLGVLVITRAHARLHLNMPHEITKRLHVNISES 240
 QY 266 GAS-----NDNHYFGYSTPMGYFDFNRFHCHSPDQRLINNNGFPRKLNLENIQV 321
 DB 241 GVGAGWQODAHQMTWTPMSGLIDANMAGWVFNPAQDLINNMTNINLVSEGEIRFVVL 300
 QY 322 KEVT---TNDGVTTIANLSTVOVSDSEYQLPYVLGSAHQGLCPFPADVMIPOYGY 378
 DB 301 KTTTESATSPPTKIYNNDLTASLAVADLTNNLTLPYPAARSETLGFYFWLPKPTQYRY 360
 QY 379 L-----TLNNGSAVGRS-----SPYCLE-YFPSSOMLRTGNNF-TFSTYFE 417
 DB 361 YLSECTRLNLPPTTGGSOQLTDSIQTGSHDIMFTYTIENAVPHILATGDEFTGIYHFD 420
 QY 418 EVFPHSSVYAHSSQSLDR-----LNNPLI--DQYLYNLRTONQSSAONKDLFSGRS 467

DB 421 TKEL--KLTHSQWQNRSLGLPPLLTEPTTEGDPHGTLPAASTRKGYQHTINNSYTEAT 478
 QY 468 PAKMSVQPKNWLPGPCYRQORVSKTIDNNNSFTMTGASKVNLNRESLIN--GTANA 525
 DB 479 ---AIRP-----AOGVYNPPMYNFEYSNGGP-----LTPVPLADT 512
 QY 526 SHKDDKDFPMSGVMIFGKESAGASNTALDNWIMDEEIKATNP-----VATERFGT 579
 DB 513 QYVDD-----PNAIFFTGYCQGLTTS-----SOELRYTFNQSKGRAPKQOFNQ 562
 QY 580 VA-VNFQSS-----TDPATG--DVHANGALP-----GVMQDRVLYLGGPIWAKIPIHT 625
 DB 563 QAEINLENTNNGTILPSPDIPGKPNHFMNLTNTYGLPLALNNTAVPVFPGQIWDKELDT 622
 QY 626 D--GHHPSPLMGPGLK-NPPQILIKTPVAPNPAEFA-TKPAFTQYSTQGVSV 681
 DB 623 DLKPRHL--VTPAPFCKNPPQGLFVKIAP--NLTDFFNADSPQOPRITYTSPFWMG 676
 QY 682 EIMELQEKNSKRNPDEVQYTSNYAKSANVDFTVDNNGLYTEPRPIG 728
 DB 677 TLFTFAKMSNNMNIQCHTT-----TAENIGYIT-PTNIG 712

RESULT 7
 COAT_PPV STANDARD; PRT; 727 AA.
 AC P04864; 065112; Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [contains: Coat protein VP2].
 OS Feline panleukopenia virus (FPV).
 OC Viruses; seRNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 ON NCB1_Taxid=10786;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85265017; PubMed=2991581;
 RA Carlson J., Rushlow K., Maxwell I., Maxwell F., Winston S., Hahn W.;
 RT "Cloning and sequence of DNA encoding structural proteins of the
 RT feline panleukopenia virus.";
 RT J. Virol. 55:574-587(1985).
 RL -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M10824; AAA47161.1; -;
 DR EMBL: M10824; AAA47162.1; -;
 DR PIR: A03701; VCPVIF.
 DR HSSP: P30129; ADPV.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 KM Coat protein: Glycoprotein.
 FT CHAIN 1 727
 FT CHAIN 1 727
 FT CARBOHYD 144 727 COAT PROTEIN VP1.
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 165 180 GLY-RICH
 SO SEQUENCE 727 AA; 80344 MW; 77B0F6E554C0C6E CRC64;
 Query Match 11.1%; Score 444.5; DB 1; Length 727;
 Best Local Similarity 23.7%; Pred. No. 2.4e-20;
 Matches 187; Conservative 117; Mismatches 295; Indels 189; Gaps 35;

```

43 RGLVLPKGYKYLGPENGLDKGPVNAADAALAEHDKAYDQOLKAGDNPYLKRNHDAEFOE 102
D 10 RGLVLPKGYKYLGPENGLDKGPVNAADAALAEHDKAYDQOLKAGDNPYLKRNHDAEFOE 69
Q 103 RLOEDTSPGNIAGRAVFOAKKRVLEPLGLVEGAKTAPGKRPEVQSPQEDSSSGIGKT 162
D 70 QTKATDMGKGIKGYHFFRAKKAIAVLDTDPDHPSTSPK-PTKRSKPHPIFINLAK- 127
Q 163 GQOPAKKRLNFGQIGDSESV-DEPPLGEPPA--TPAAVGTPTMASGGGAMADNNE 216
D 128 -----KKKAGAGQYKRDQAPMSDGAQVDPDQAPVNERATGSGNGSGGGG-----GG 176
Q 217 GADGVGNASGNW-----CDSTWLDGRVITTSRTWALPTYNHLYKQI-----SS 262
D 177 GSGGVGISTGTFNNQTEKFLENGW--ETANSRLVHLMPSEENYKRVVANNMOKTA 234
Q 263 ASTGASNDNHYFGYSTWGYEDFNRFHCFSPRDMORLNNMGFRPRKLNFKLENIOVK 322
D 235 VKGNMALLDTHVQIVTTPMSLVDANAMGWFMNPGMQLIVNTMSELHVSFEQEIFNVVLK 294
Q 323 EVT---TNDGVTTANNITSTVOVPSSEYOLPYVLSAHGCL-----PPPA----- 368
D 295 TVSESATOPPTKYVNNDLTASIMVALDSNNMPTPAMRSETLGFYPMKFTIPTPMRY 354
Q 369 ---DVFMIPOYGYLTLNNGSQAVGRS-----SFYCLE-YFPQOMLTGNN--TFS 413
D 355 FQWRTILPSH-----TGISGTPNINHGTDPPDVQPTTENSVPVHLATGDFPATGT 408
Q 414 YTFEEVPHSSVAHSQSLDRLMN--PLIDQYLYLNRTONOSGSAONKDLLFSRSPAGM 471
D 409 FFFDOCKP--CRLTHTWQTRALGLP-----FLNSLPQEGATNFGDIGVQDCKRGV 459
Q 472 S-----VQPKN-WLPGRCYQQRVSK-----TKDNNNSNTWT 504
D 460 TOMGNITVITATITMRPAVEYSAVYSFEASTQCPKIPITAGRGGAQTDENQAA--D 516
Q 505 GASKY---NLNGRESIINPGT-----AMASHKODEDKFFPMSCWMIFGKESAGASNTALD- 556
D 517 GDPRIYAPGROHGOCTTTGETPERFTYIAHODT-----GRYPAGDMIQINIF 563
Q 557 NVMITDEEIKATNPVALTERFGTVAVNFQSSSTDPATGVHMGALPGMWQDRVYLQ 616
D 564 NLPTNDVILLPTDPIG---GKTGINY--TNIENVYGLTALANNP-----PYVNG 610
Q 617 PIAKTIPTDGHFHPSPLMGFGK-----NPPQILIKTPTVPAN--PPA 660
D 611 QIMKEEPTD-----LKPRLHVAAPVCONNCPQOLFVKVAPULTNEYDPA 657
Q 661 EFSATKFAFITQYSTGVSEIEMWLOKENSKEKNPEVOYTSYVAKSANDFTVDNNG 720
D 658 SANMSR---LVYTSDFWKKGLVFKALRASHHTNPIQOMKIN-----VDQFN 703
Q 721 YTEPRPIG 728
D 704 YL-ENNIG 710

```

```

RT "Porcine parvovirus: DNA sequence and genome organization.";
RL J. Gen. Virol. 70:2541-2553 (1989).
CC -1- SUBUNIT: NATURE VIRION CONTAINING THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D00623; BAA0502.1; -.
DR PIR, B33302; VCEVPP.
DR HSP, P30129; 4DPV.
DR InterPro: IPR001403; Parvo coat.
DR Pfam: PF00740; Parvo coat; 1.
KM Coat protein, glycoprotein.
FT CHAIN 1 749 COAT PROTEIN VP1.
FT CHAIN 1 749 COAT PROTEIN VP2.
FT DOMAIN 195 210 POLY-GLY.
FT CARBOHYD 192 192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 671 671 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 749 AA; 82872 MW; 6CA13CB97B93418 CRC64;
Query Match 11.1%; Score 441.5; DB 1; Length 749;
Best Local Similarity 25.3%; Pred. No. 3.8e-20;
Matches 195; Conservative 109; Mismatches 311; Indels 157; Gaps 37;
44 GLVLPKGYKYLGPENGLDKGPVNAADAALAEHDKAYDQOLKAGDNPYLKRNHDAEFOE 103
31 GLTLPFG-TILGPNNSLDQGEPTPNSDAAAKEHDAVKIKSKNPTFYSADEKFIKE 89
104 LOEDTSPGNIAGRAVFOAKKRVLEPLGLVEGAKTAPGKRPEVQSPQEDSSSGIGK-- 161
90 TEHAKDYGKIGYHFFRAKKAIAVLDTDPDHPSTSPK-PTKRSKPHPIFINLAK- 146
162 -----TGQOPAKKRLNFGQIGDSESVDPQPLGEPPATP-AAVGTPTMASGGGAPMA 212
147 APHPIFINLAKKAKGTSNNSMSBNVQHNPIIN--AATELSATGNESGGGGG--- 201
213 DNNEGADGVGNASGNW-----CDSTWLDG--RVTTSTRTWALPTYNHLYKQI-----SS 262
202 -GGAGAGGVGVSTGSEFNQTEFQYLGELVRIYAHASRLIHLMPHEITYKRIHVINSES 260
263 ASTGAS-NDNHYFGYSTWGYEDFNRFHCFSPRDMORLNNMGFRPRKLNFKLENIOVK 321
261 GSAQGVQDDAHQOMTTPMSLIDANAMGWFMNPGMQLIVNTMSELHVSFEQAIRNVVL 320
322 KEVT---TNDGVTTANNITSTVOVPSSEYOLPYVLSAHGCLPPRPADVEMIQGY 378
321 KITTESATSPPTKIYNNDLTASIMVALDNTNNTLPYTPAARSETLGFYPLPKPQYRY 380
379 L-----TLNNGSQAVGRS-----SFYCLE-YFPQOMLTGNNF-TFSYTFE 417
381 YLSCIRLNLPPTTGSQOQTDSIQGLHSDIMFYTIENAVPIHLRTGDFEFTGYHDF 440
418 EVFPHSSVAHSQSLDR-----LNNPLI--DOYLYLNRTONOSGSAONKDLLFSRGS 467
441 TKPL--KLTHSWQTNLSGLRPKLLTEPTTEGQOHGCTLPAASTRKGYHOTINNSTEAT 498
468 PAGMSVQPKMWLPGPCYRQQRVSKTYTDNNNSNFTWTGASKYNNLNGRESIINP--GTAMA 525

```

```

Db      499  ---AIRP-----AQGVNTPYNNFEISNGPF-----LTFIVETADT 532
Qy      526  SHKDEDKFPFMSGVWIFGKESAGASNTALD--NVMITDEEIK--AIRP-----VAT 574
Db      533  QYNDEPN--NF-----GALRPTMDVOHGHLTTYSQLELEYTFNPOSCKGRAPK 577
Qy      575  EREFGTVA--VNFOSS-----TDPATG--DVHANGALP-----GHWQODRDVYLQGPWA 620
Db      578  QQFNQOAPLNTLNTNNGTLLPBDPFIGCKSNMHEPMTLNTVYGPPLTALNTNTPAVPNCQIWD 637
Qy      621  KIPHTD--GHFHPSPILMGFGFLK--NEPPQILINTEVPANPAPFSA--TKFAFITYOYST 676
Db      638  KELDIDLKPRLLH--VTAPFVCQNNPNPGLFLVLIAP--NLDDFNADSPQQRILITYSN 691
Qy      677  QGVSVFEIHELQKENSKRWNPEVOYQTSYNAKSAANDPYDNGCLYTEPRPIG 728
Db      692  FWWKCKTLFTTAKORSSNMNPIQOHTT-----TKENICKYI--PTNIG 732

```

RESULT 9	COAT_MEVA	COAT_MEVA	STANDARD;	PRT;	722 AA.
ID	COAT_MEVA				
AC	P27437				
DT	01-AUG-1992	(Rel. 23, Created)			
DT	01-FEB-1996	(Rel. 33, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Coat protein VP1 [Contains: Coat protein VP2].				
OS	Mink enteritis virus (strain Aashviri) (MEV).				
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.				
NCBI_TaxID=10793;					

RP SEQUENCE FROM N.A.
RX MEDLINE=91202123; PubMed=2016597;
RA Katsunemari T., Horiuchi M., Hama E., Yaguchi K., Ishiguro N.,
RA Goto H., Shimagawa M.;
RT "Construction and nucleotide sequence analysis of an infectious DNA
RT clone of the autonomous parvovirus, mink enteritis virus.";
RL J. Gen. Virol. 72:867-875(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	PIR; B38350; VCPVME.				
DR	HSSP; P30129; 4BPV.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
KM	Coat protein; Glycoprotein.				
FT	CHAIN	1	722	COAT PROTEIN VP1.	
FT		139	722	COAT PROTEIN VP2.	
FT	DOMAIN	160	177	GLY-RICH.	
FT	CARBOHYD	163	163	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	185	185	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	202	202	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	318	318	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	581	581	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	643	643	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	655	655	N-LINKED (GLCNAC. . .)	(POTENTIAL).
EQ	SEQUENCE	722 AA;	79823 MW;	99DDBCAB2EF9F622 CPG64;	

```

Query Match      11.0%; Score 437.5; DB 1; Length 722;
Best Local Similarity 23.6%; Pred. No. 6.5e-20;
Matches 186; Conservative 117; Pident 295; Indels 189; Gaps 35;

QY      44  GLVLPGYLYLPFNGLDKGEPPVNAADAALAEHDKAYDQQLKAGNDPYLYRYNHADAEPFER 103
      ||| ||||| ||| ||: ||| | : ||| ||| ||| : ||| ||| : ||| : ||| : |||

```

```

Db 6 GLVPRGKYLIGPBNLSIDJSEPTNPBDAAKHDEHYAALYLRSGNPVLYTSPADQRIIDQ 65
Qy 104 LOEDTSFGNGLGRAVFOAKKRVLEBJGVERGAKTAPKKRPVEOSPOEDPSSSIGIKTG 163
Db 66 TKOATDMGKLGHYFPRAKKAIAVLUTDTPHPSRPTK--PTKRSKPPPHIIFINLAK-- 122
Qy 164 QOPAKKRLNFGOTGDESEV-----DPOPLGEPPA--TPAAVGPPTTMASSGGGAPMADNEG 217
Db 123 ----KKKAGAGOVKRDNLAPMSDGAIVDPGGQPAVRNERATGSGNGSGGGG-----GGG 172
Qy 218 ADGVSNAAGSNWH-----CDSITWJGDKVITTTSTFTALPPTNNHLYKOT-----SSA 263
Db 173 SGGVIGISTGFENNQGEFKPLENGWY--BITANSSVLHAPNBEENKYKRVVNNMDXTAV 230
Qy 264 STGASDNHNYFGVSTPMGWYFDENRPHCHSPSPDMQORLLNNNMGRRPKRLNFKLEIOVKE 323
Db 231 KGNMALDDTHVOQVTFPMSLVDNANAGWFWNFGDMOLVNTMSSEIHVSFEOELFNVALKT 230
Qy 324 VT---TNDGVTTIANLSTVOVFPDSEYOLPYVLSGAHQCL-----PPFPA----- 368
Db 291 VSSSAIQPPTKYVNDLTASLVALDASNNTMPTPAAARSTLGFYPMKPTIPIPMKRYTF 350
Qy 369 --DVPMIPOYGYLLTNNGSQAVGRS-----SFYCLE--YPPSQMLRTGNNF--TFSY 414
Db 351 QMDRTLIPSH-----TGTSGETPIVIYGTDPDDVQFTIENSVPHLLRTGDEFAITGF 404
Qy 415 TEEVFPFHSSYAHOSLDRLLM--PLIDQVLYLNRTONOSGSAQNKDILFSRGSFAGMS 472
Db 405 FFDCKP--CRLTHTWQTRNLRALP--FLNLSLPOSGAGTNFEDIQVQDCKRGVTF 455
Qy 473 -----VQPKN--MLRPECYRQORVSK-----TKTDNNNSNFWTWG 505
Db 456 QMGNTDYITEATITNPAEYVGSAPYFSFESTQGEFKTPIAAGRGAGQOTENQAA--DG 512
Qy 506 ASKY--NINGRESIINPGT--AMASHKDEDEKFFMSSGVMIFGESAGASATLAD-N 557
Db 513 DRYAAGROHQKQKTTTGTETPERFTYIAHDT-----GKYPADMIQINFN 559
Qy 558 VMTIDEELIKATNPVATERFGTVANFOSSSTDPAIGDVMHMGALPGMWQODRDVYLQGP 617
Db 560 LPTNNNVLLPDPDIG-----GKTGINY--TNIPTNYGTLTANNVP-----PYVNGQ 606
Qy 618 IWAKIPIHTDGHFHPSPFLMGFGLK-----NPPQILLIKOTPVYPAV---PPAE 661
Db 607 IMDKERTD-----LKRLHVNAPFVQCQNNCPGQLPVKVAIPMLTMEYDAS 653
Qy 662 FSAITKASFTIOTSGOVSVETEMELQKENSKRANPEVOYTSNYAKASANDFTYDNNGLY 721
Db 654 ANMSR-----IVYTSDEPMWKGKLVFPAKRLASHTNWPIQMSIN-----VDNQFNY 699
Qy 722 TEPRPIG 728
Db 700 L-PNNIG 705

```

	RESULT	10	
	COAT_FPV19		
ID	_COAT_FPV19	STANDARD;	PRT; 727 AA.
AC	P484f0;		
DT	01-MAR-1992	(Rel. 21,	Created)
DT	01-MAR-1992	(Rel. 21,	Last sequence update)
DT	16-OCT-2001	(Rel. 40,	Last annotation update)
DE	Coat protein VP1 [Comments: Coat protein VP2].		
OS	Feline panleukopenia virus (strain 193) (FPV).		
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.		
OX	NCBI_TaxID=10787;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=193/70;		
RX	MEDLINE=91073139; PubMed=2174965;		
RA	Martyn J.C., Davidson B.E., Studdert M.J.;		
RT	"Nucleotide sequence of feline panleukopenia virus: comparison with canine parvovirus identifies host-specific differences."		

RL J. Gen. Virol. 71:2747-2753 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CU-4;
 RX MEDLINE=91272479; PubMed=1647068;
 RA Parish C.R.;
 RT "Mapping specific functions in the capsid structure of canine
 parvovirus and feline panleukopenia virus using infectious plasmid
 clones";
 RL Virol. 183:195-205 (1991).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55115; CA38911.1; -
 DR EMBL; M38246; AAC37928.1; -
 DR EMBL; M38246; AAC37929.1; -
 DR PIR; B36608; VCPVFP.
 DR HSSP; P30129; ADPV.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat_1.
 DR Coac Protein; Glycoprotein.
 KW CHAIN 1 727
 FT DOMAIN 144 727 COAT PROTEIN VP1.
 FT CARBOHYD 165 190 GLY-RICH.
 FT CARBOHYD 168 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 165 180 GLY-RICH.
 SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621FF5 CRC64;
 Query Match 10.8%; Score 431.5; DB 1; Length 727;
 Best Local Similarity 23.5%; Pred. No. 1.5e-19;
 Matches 185; Conservative 117; Mismatches 297; Indels 189; Gaps 35;
 QY 43 RGLVLPFGKYLGPENGLDKSPVNAADAALEHKKAYOOLKAGDNPTRLNHDAERQ 102
 DB 10 RGLVLPFGKYLGPENGLDKSPVNAADAALEHKKAYOOLKAGDNPTRLNHDAERQ 69
 QY 103 RLODSTFGNGLGRAVFOAKKRVLEPLGLVEGAKTAPGKRPVEOSPOEDSSSGIGKT 162
 DB 70 QTKAKOMGKIGHYFPKAKAIAPVLTDPDHSSTPK-PIKRSKPPHITINLAK- 127
 QY 163 GQCPAKRLNFGQGTDSSEVP---DPOPLGEPPA--TPAAVGPPTMASGGGAPWADNNE 216
 DB 128 -----KKKAGAGQYRDLAPWSDAVQPDGQPAVRERAKVSGSGSGGG-----GG 176
 QY 217 GADVGNAAGNWH-----CDSTWLGSRVITSTRTALPTNNHLYKQI-----SS 262
 DB 177 GSGVGISTGTFFNNQTEKFLENGWV--EITANSRLVHLMPESENKRVAVNMMDKTA 234
 QY 263 ASTGASNDNHFGYSTPGVGFDFNRFHCHFSPRDQRLINNNGWFRPKLFLKLFENIOVK 322
 DB 235 VKGNMADDIDIVQVITPMSLVDAANAGWENFGDQILVNTMSLHLVSEFOELFNVYLK 294
 QY 323 EVT---TNDGVTTIANLITSTVOVFSDEYOLPYVLGSAHQ3CI-----PPPPA----- 368
 DB 295 TVSSASATQPTPKVYNNNDLTASLWALDSNNTMPTPAAMRSGLTGFYFWKKTITPTWRY 354
 QY 369 ---DVFMTPQYGYLTUNGSAVGRS-----SYCYE-YFPSMLATGNPF-TFS 413
 DB 355 FQMDRTILPISH-----TGTSGTPTNVYHGTDPDVOGYTLENSVPHALLATGDEFATGT 408

QY 414 YTFEEVFFHSSVAHSQSLDLAMN--PLIDQYLYLNRTQSGSAQNKDLFFSGSPAGM 471
 DB 409 FPFDCRP--CRULTQGTNRALGLP-----FLNSLPSEBQATNFGDIGNVODRRGV 459
 QY 472 S-----VOPKN-WLPGCYRQGRVK-----TKTNNNSNFTWT 504
 DB 460 TQMGNTDYITREATIMBPAEYGSAPYSPFASVQGFPTPIAAGRGACQDEMOAA---D 516
 QY 505 GASKY---NUNGRESIINPCT---AMASHKDEDEKFFPMSGWMIGKESAGASNTALD- 556
 DB 517 GDDPYAFGRHQGTQTTTGTGPERFTYIAHQT-----GRYPEGMIQNIWF 563
 QY 557 NWMTDEEIKATNPVATERFGTAVVFOSSSTDPAQGVHANGALPGWVQDRDYVLOG 616
 DB 564 NLQVTDNDVLLPDPDG---GKMGINY--TNFNNTGTLTALNNP-----PYVPG 610
 QY 617 PIMAKIPHTDGHFHPSPLMGFGGLK-----NPPQILIKNTVVPAN---PPA 660
 DB 611 QIMDKERTD-----LKPRLHYNABFVCCNNCPGQLFVFAENLTNEYDPDA 657
 QY 661 EFGATKFASTIQYSTQGVSEIEMELQKNSKRNPEVQYTSNVAKSANVDFTVDNGL 720
 DB 658 SANMSR---ITVTSDFWKKGLVFPAKLRASHTMPIQOMSTN-----VDNQFN 703
 QY 721 YTEPRPIG 728
 DB 704 YV-PNNIG 710
 RESULT 11
 COAT PAVCB STANDARD; PRT; 727 AA.
 ID COAT PAVCB
 AC 011213;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2].
 OS Canine parvovirus (strain B) (CPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=59284;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91272479; PubMed=1647068;
 RA Parish C.R.;
 RT "Mapping specific functions in the capsid structure of canine
 parvovirus and feline panleukopenia virus using infectious plasmid
 clones";
 RL Virol. 183:195-205 (1991).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 OF A COMBINATION OF VP2, VP3, AND SOME VP1.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M38245; AAB02799.1; -
 DR EMBL; M38245; AAB02800.1; -
 DR HSSP; P30129; ADPV.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat_1.
 DR Coac Protein; Glycoprotein.
 KW CHAIN 1 727
 FT CHAIN 144 727 COAT PROTEIN VP1.
 FT CARBOHYD 168 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 727 AA; 80342 MW; 115FELAL79098EBE CRC64;
 Query Match 10.8%; Score 430.5; DB 1; Length 727;
 Best Local Similarity 23.2%; Pred. No. 1.8e-19;
 Matches 184; Conservative 112; Mismatches 297; Indels 201; Gaps 32;
 QY 43 RGLVLPQKYLPGNGLDKGEPPVNAADAALAEHKAVDQOLKAGDNVYLNNHDAEFOE 102
 DB 10 RGLVLPQKYLPGNGLDKGEPPVNAADAALAEHKAVDQOLKAGDNVYLNNHDAEFOE 69
 QY 103 RLOEDTSFGNLTGRAVPAKRRVLEPLGLVEBGAKTAPGKKRPVPGSPQEPDSSSGIGKT 162
 DB 70 QTDKADKMGKIGHYFPRAKKAIAVLDPDHDSTSRPRK-PIKRSKPPHIFINLAK- 127
 QY 163 GQDPAKRLNFGQGTGSESV- -DPOIGEPPA-TPAAVGETTMASSGGAPMADNNE 216
 DB 128 - - - - -KKKAGAGGVKRDNLAPMSDGAIVPDGGQPAVNERBATSGNSGGGG- - - - -GG 176
 QY 217 GADGVNAGSNM- - - - -CDSTWLGDRVITTSRTWALPTNNHLYKQI- - - - -SS 262
 DB 177 GSGGVGISTGFENNQTFKLENGWV- -ETANSSRLVHLMPESENYRVVNNMDKTA 234
 QY 263 AATGASNDNHFGYSTWGYFDFNRFFCHSPRDWQRLINNNGFRPRKLNFKLFNIQVK 322
 DB 235 VNGMMALDDIHAQVTFPMSLVANAMGWFRNPGMQLIVTMSLHLVSEGELEFNVLK 294
 QY 323 EVT- - -TNDGVTTIANLSTVOVSSEYQLPYVLGSAHQGL- - - - -PPPPA- - - - - 368
 DB 295 TVSSAATQPPTKVNNDLTASLWALDLSNNTMPTPAAMSEETLGFYPMKPTLPTPRY 354
 QY 369 - - -DVFMIPOGYVTLNNGSAVGRS- - - - -SFCLE-YFPGMRTGNF-TFS 413
 DB 355 PQMDRTILP- - - - -TGTSPTNTNYHGTDPDVOFYTIENSVPPLHRTGDEFYGT 408
 QY 414 YTEEEVFFSHSYAHOSGLDRIMN- -PLIDQVLYLNTQNSGSAQNKDLFRRGSPAG 471
 DB 409 FFPDCKP- -GLTHTWQTNALGLP- - - - -FLNLPSEGAATNGDI- - - - -GV 451
 QY 472 SVQPKMLPQPCYRQGVSKTKTDNNNSN- - - - -TWGASKYNLNGRESIINP- 520
 DB 452 - - - - -QDDKRGVTOGMGNTNYITEATIRPAEVSAPAYSFEEASTQPFK 497
 QY 521 - - -GTAMASHKDEDEKFFPMGWIHFKESAGASNTLD- - - - - 556
 DB 498 TPIAAGRGGAQTDEMQADGNPRYAFGRHQKTTTGETPERPTYIAHODTGRYPEGDW 557
 QY 557 - - - - -NMTDEEELKATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDR 610
 DB 558 IQNINRPLPTNDNVLLPTDPIG- - -GKTGINV- - -ININITYGPLTALANVP- 604
 QY 611 DVTLOGPIMAKIPTDGHFHPSPLMGFGK- - - - -NPPQILIKNTTPVAN 657
 DB 605 PVPVNGQIMKEPDTD- - - - -LKPRLVHNAVPVCONCNGQFLVKAAPLNTN 651
 QY 658 - - - - -PPAEFSATKASFITQYSTQVSVEIMELQKSKRNNEVQTSYASANDFT 714
 DB 652 EYDPDASANSR- - - - -LVYTSDFWKGKLVKALRASHTWNPLOQMSIN- - - - - 697
 QY 715 VDNNGLYTEPRPIG 728
 DB 698 VDNQFNV-PSNIG 710

RESULT 12
 COAT MIMIM STANDARD; PRT; 718 AA.
 ID COAT MIMIM
 AC P07302; Q9NMH3; Q9NMH2;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 (Contains: Coat protein VP2).

OS Murine minute virus (strain MMV) (Murine parvovirus).
 OC Viruses; ssRNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=107955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=66115415; PubMed=3502703;
 RA Astell C.R., Gardner B.M., Tattersall P.;
 RT "DNA sequence of the lymphotropic variant of minute virus of mice,
 MMV(1), and comparison with the DNA sequence of the fibrocytic
 prototype strain.";
 RL J. Virol. 57:656-669 (1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85242059; PubMed=3855242;
 RA Shih R., McMaster G.K., Hirt B.;
 RT "DNA sequence comparison between two tissue-specific variants of the
 autonomous parvovirus, minute virus of mice.";
 RL Nucleic Acids Res. 13:3617-3633 (1985).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718
 RA Llamas-Saiz A.L., Agbandje-McKenna M., Winkler W.R., Bratton J.,
 RA Tattersall P., Rosemann M.G.;
 RT "Structure determination of Minute Virus of mice.";
 RL Acta Crystallogr. D 53:93-100 (1997).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X02481; CAB46507.1; -
 CC EMBL; X02481; CAB46508.1; -
 DR EMBL; M12032; AAB69569.1; ALT_INIT.
 DR PIR; B23008; VCPVIM.
 DR PDB; 1MTM; 25-FEB-98.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein; 3D-structure.
 FT CHAIN 1 718
 FT CARBOHYD 132 718 COAT PROTEIN VP1.
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 502 502 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 157 172 GLY-RICH
 FT CONFICT 144 144 A->G (IN REF. 2).
 SQ SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;
 Query Match 10.6%; Score 424.5; DB 1; Length 718;
 Best Local Similarity 23.7%; Pred. No. 4.1e-19;
 Matches 184; Conservative 104; Mismatches 298; Indels 199; Gaps 31;
 QY 45 LVLPGYKYLPGNGLDKGEPPVNAADAALAEHKAVDQOLKAGDNVYLNNHDAEFOEL 104
 DB 1 LVLPGYKYLPGNGLDKGEPPVNAADAALAEHKAVDQOLKAGDNVYLNNHDAEFOEL 60
 QY 105 QEDTSFGNLTGRAVPAKRRVLEPLGLVEBGAKTAPGKKRPVPGSPQEPDSSSGIGKTG- 163
 DB 61 KDAKMGKGVGHYFPRTKRAFAPKLATDSE- - - - -PG- - - - -TSGVSRACK 101
 QY 164 - - - - -QDPAKRLNFG- - -QTDSESVPPDPLG- - - - -EPPATPAAVGPTTM 203
 DB 102 RRRPRAVITINAPRAKKTLSAQAQSSQYTMDSQPPGGAANAARAVEAALDQPG- - 159
 QY 204 AGGGAPMADNNEGADGVNAGSNMHCDS- -TWLGD- - -RVITTSRTWALPTNNHLYK 258
 DB 160 GSGGG- - - - -GSGGGGVGVSTGSDNCTHYFLADGWETVALATRLVHLMPESENYC 213

QY	25	QI-----SSASTGASNNHHYGVSTPMGYEDFNRFCHEFSRPMQGLINNMGSPRKR	311
Db	214	RIRVHHNTDTSVGVGNMAKDHAHQIWTPTSLVDANAMGWLSPSMQOICNTMSQMLVS	273
QY	312	LNFKLFNIOQKEVTIND---GVTTIANNLSTVOFSDESYQLPYLGSAHQCLPFP	367
Db	274	LDOEIFVVLVKTYTEODSGGOAIKIKYNNDLTTCMMVAADSNIIIPYTPAANSMETLGFYP	333
QY	368	ADVFMIPQGY-----LTLNN-----GSAQVGHSPCYCLEYPS--QMLRTG	407
Db	334	WKRTIASPRHYCYCDNRDLSVTYENOGTIEHNHWGTPKGMISQFTIENTQOQLTLRTG	393
QY	408	NNF--TFSYTFEEVPFHSSYAHSGSLDRLN--PLIDQYLXLYNRTONOGSQAOKDLIFS	466
Db	394	DEPATGYUYEDTNPV--KLTHWTQTRQLGQPLPLSTF---PEADTDRGT-----LTA	444
QY	465	RGSBAGMSVQPKWL-----PGCYAQQRVSKTKTDNNNSN	500
Db	442	QGSBHGTQMEVNWVSEAIRTRPAOVGFQCPHNDEPASHG--FAAPKVPADVTQGDRE	500
QY	501	FTVTGASKYNLNGRESIINPGTAMASHKDEDKF-----FPMGVMIFGKSAGAS	551
Db	501	ANGSVRSYKQGHENWAAHGAPREXYTWDENFGSGRTRDGFIOAPLVPPINGIL	566
QY	552	NTPLDNNMTIDDEBEIKATNPVALTERGTYAANFGS--STDPATGDVAMGALGGMWQD	609
Db	561	TNA-----NFIGTKN---DIHSNVSNGYGLTAFSH-----P	599
QY	610	RDVYLOGPIWAK---IPHTDGHFHPSPLMGFGGLKNRPQOILIKNTP--VPANPPEAFSA	664
Db	591	SPVYPQOQIDKRELDLHKKRPLHITAPV---CKNNAFGOMLVRLGPNMLDQYDNGATL	644
QY	665	TKRASPTQYSTQGVQVEIEMELQKNSKRNPMPEYQTSNRYAKSANDPYTDNNG	719
Db	648	SRIVYGTGFPMWKKLTKRAKLRA---NTTAMPVVQ-----VSDVDNG	686

```

RESULT 13
COAT_PAVCN
ID_COAT_PAVCN STANDARD; PRT; 748 AA.
AC P12930; O84387;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2].
OS Canine parvovirus (strain N) (CPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
ON NCBI_TaxID=10791;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86062992; PubMed=2824850;
RA Reed A.P., Jones E.V., Miller T.J.;
RT "Nucleotide sequence and genome organization of canine parvovirus ";
RL J. Virol. 62:265-276(1988).
-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
OF A COMBINATION OF VP2, VP3, AND SOME VP1.
-1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC	EMBL	M19296	AA67460.1	-
DR	EMBL	M19296	AA67461.1	-
DR	PIR	B29962	VCPVCP	
DR	HSSP	P30129	4BPV	
DR	InterPro	IPR001403	Parvo_coat	
DR	Pfam	PF00740	Parvo_coat	1
DR	Coat protein	Glycoprotein		
kw				

FT	CHAIN	1	748	COAT PROTEIN VP1.
FT	CHAIN	165	748	COAT PROTEIN VP2.
FT	CARBOHYD	189	189	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	228	228	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	344	344	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	607	607	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	669	669	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	681	681	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	DOMAIN	186	203	GLY-RICH.
FT	CONFLICT	27	27	M -> XISM (IN REF. 1: AAA67460).
SO	SEQUENCE	748 AA;	82715 MM;	13E487C6862C73B6 CMC64;

Query Match	10.6%;	Score 421.5;	DB 1;	Length 748;
Best Local Similarity	23.0%;	Pred. No. 6.7e-19;		
Matches 184;	Conservative 114;	Mismatches 287;	Indels 215;	Gaps 34;

Qy	44	GLVLPFGXYKYLFPFSGDLKGBEVNADAAALHEDRAVDOQLKAGDNPFYLRYNHADAEPQBR	103
Db	32	GLVLPFGXYKYLFPFSGDLKGBEVNADAAALHEDRAVDOQLKAGDNPFYLRYNHADAEPQBR	91
Qy	104	LQEDTSGNLRGAVFOAKKRVLEPLGLVEGATAPAEKKRPVEQSPQEPDSSSGIGTKT	163
Db	92	TKDAXMDGCKIGHFFRAKKAIAEVLTDTPHSPSTREPTX-PIYKSKPPHIFINLAK--	148
Qy	164	QQPAKKRLNFGQTDSSVSP-----DPOPLGEPRA--TPAAAGPPTMAAGGGAPMADNNEG	217
Db	149	-----KKKAGAGQVYKRDNLAPMSDGAQVDPDGGQPAVRNERATGSGNGSGGG-----GGG	198
Qy	218	ADGVGNASGNWH-----CDSTWLGDRVITTTSTRTMALPTYNHLYKQI-----SSA	263
Db	199	SGGVGISTGTETNNQTEPFLENGVW-ETIANSRLVHLNMPESBENYRVRVYVNNMDKTAV	256
Qy	264	STGASNDNNHFGYSTPMGYDPFENFCHFSRDMQORLLNNWGFPPKRLNPLFLIOYKE	323
Db	257	NGNNALDDIHAQIVTPMSLYDANAMGWENFGDQOLVINTSECHLVSEQDIEIFNVILKT	316
Qy	324	VT-----TNDGVYTLINNLSTVQVESDSEYOLPYLYGSAHQCL-----PEPPA-----	368
Db	317	VSEBATOPEPTKYVNNDLTASLMAVLDSNNMPTPPAARSETLGYPMKPTIPIYRWRYF	376
Qy	369	--DVFMIPQYGYLLTNNGSQAVGS-----SFYCLB-YEPSQMLRTGNNF--TFSY	414
Db	377	QMDRTLLPSH-----TGTSQTPNIIYHGTDPPDVQFYTIENSFVHLLRTGDDEBATGWF	430
Qy	415	TFEVFPHSSSTAHOSQIDRLMN--PLDQVLYLINRTQONGSAGNKKLLRSRGSBAGMS	472
Db	431	FFDCKP--CRLTHWQJNTRALGLRP-----FLNSLPQSEBACATNFGI-----GV--	472
Qy	473	VQPKWMLPGFCYRQOARVSKTLDNNNSNF-----TWGASKYNNNGHESIINP--GT	522
Db	473	-----QQDKRGRVTOGNGNTYITTEATIMRAEYGYAPYSSPEASTGCPFKT	519
Qy	523	AMASHK-----DDEDKXFFPMGCVMI FGESAGASNTALD-----	556
Db	520	PIAAGRGGAQYENQADADGDPRY-----AFGRQHGQKTTTGTGTPRFTYIAHQDTGR	572
Qy	557	-----NMWIDEBEIKATNPVATRFQVVAANFOSSSSTPATGDVHAGALPG	604
Db	573	YPSBGDWIQNINENFLPVTNDVNLPLTDPJG---GKTGINY--TNLFNTYGPILTALNNP-	625
Qy	605	MWQDRDVLYLGPITWAKI PHTDGHFHSPRLMGGFGIK-----NBPQILIKN	651
Db	626	-----PYPRPGQIMDKPEFTD-----LKRLHVNAPFVCONNPGQLFVAV	666
Qy	652	TPVPAV--PPAEFSATKEASFITQYSTGVQSVLEIEMLOKENSGRANPEVUYTSNYAKS	708
Db	667	APNLITNEYDPPASANMSR-----IYTSDFMWKXGLVFRAKULRASHTWNPICQMSIN----	718
Qy	709	ANVDPTDNNGLYTEPRPIG 728	
Db	719	-----VDNQENYV-PSNIG 731	

```

RESULT 14
COAT_PAVCD STANDARD; PRT; 737 AA.
ID COAT_PAVCD STANDARD; PRT; 737 AA.
AC P17455;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Canine parvovirus (strain CPV-2 Cornell 320) (CPV).
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=10790;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89020796; PubMed=3176341;
RA Parvovirus C.R., Aquadro C.F., Carmichael L.E.;
RT "Canine host range and a specific epitope map along with variant
RT sequences in the capsid protein gene of canine parvovirus and related
RT feline, mink, and raccoon parvoviruses."
RL Virology 166:293-307(1988).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL_M33255; AAA47158.1; -.
DR PIR; A31163; VCPVCD.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 737 COAT PROTEIN VP1.
FT CARBOHYD 154 737 COAT PROTEIN VP2.
FT CARBOHYD 178 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 175 192 GLY-RICH.
SQ SEQUENCE 737 AA; 81504 MW; 39FEB1D3E5435EC9 CRC64;
Query Match 10.2%; Score 407.5; DB 1; Length 737;
Best Local Similarity 22.8%; Pred. No. 4.9e-18;
Matches 181; Conservative 112; Mismatches 301; Indels 201; Gaps 32;

```

```

322 KEVT---TNDGVTIANLSTVQVSESEYOLPVYLGASHOGL-----PPFPA----- 368
DB 304 KYVESATOPPTKRVYNNDLTASLMAVLDSSNNTWFFPAPAMRSETLFFYWKPTIPTPMKY 363
QY 369 ---DVFMIPQGYLYTLNNGSQAVGRS-----SFYCLE-YPPSQMLRTGNF-TF 412
DB 364 YFGWDRLLPSH-----TGTSGTPTNIYHGTDPDVOGYTLIENSVPVHLRTGDEPATG 417
QY 413 STTFEEVPPHSSYASHOSLDRLM--PLIDQYLYINRQONSGSAGNDLFRSSSPAG 470
DB 418 TFEFDCKP--CRULTWQTRALGLP-----PLNSLPQSGATNPGDI-----G 460
QY 471 MSVQPGMWLPGPCYQQRVSKTKTDNNNSNF-----TWGASKYNLNGRESIIP- 520
DB 461 V-----QDQKRGVYQMGNTNYITTEATINPAEYGSAPYSPFASITQGP 506
QY 521 ---GTAMASHKDEDEKFPMSGVMIFGESAGASNTALD----- 556
DB 507 KPIAAGRGGAQTDENQADGNRVAFGHQGKTTTGTETPERFYIAHQDTGRYPEGD 566
QY 557 -----NMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAMGALPGWQD 609
DB 567 WIGINIFNLPTVINDVLTLPDPIG---GKTGINY--TWIPTYGPTALNNVP----- 614
QY 610 RDVYLGPIWAKIPHTDGHFHPSPLMGFGGLK-----NPPQIILKNTPPVA 656
DB 615 -PYTPNGQIMDEXEDTD-----LKRHLVNAAPVCCNNCPQGLFYKVAANLT 660
QY 657 N--PRAESATKFAFITQVSTQVSEIEMELQKNSKRNPEVQYTSNYAKSANVD 713
DB 661 NEYDPDASANMS---ITYSDPFMWKGLVFAKAKARASHNTNPIQMSIN----- 707
QY 714 TYDNNGLYTEPRPIG 728
DB 708 -VDNQFNYY-PSNIG 720

```

```

RESULT 15
COAT_PAVHH STANDARD; PRT; 722 AA.
ID COAT_PAVHH STANDARD; PRT; 722 AA.
AC P03136;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Hamster parvovirus H1.
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=10799;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83112183; PubMed=6823009;
RA Rhode S.L. III, Paradiso P.R.;
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its
RT genes by hybrid-arrested translation."
RL J. Virol. 45:173-184(1983).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X01457; CAB57285.1; ALT_SEQ.
DR PIR; A03699; VCPV2.
DR HSSP; P07302; 1MW.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 722 COAT PROTEIN VP1.

```

FT	CHAIN	131	722	COAT PROTEIN VP2.	(POTENTIAL).
FT	CARBOHYD	178	178	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	218	218	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	647	647	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	675	675	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	155	170	GLY-RICH.	
SO	SEQUENCE	722 AA;	79737 MW;	6B8678391A5DC31	CRG64;

Query Match	10.1%;	Score	404;	DB 1;	length	722;			
Best Local Similarity	23.3%;	Pred. No.	7.9e-18;						
Matches	175;	Conservative	108;	Mismatches	290;	Indels	178;	Gaps	30;

```

0Y 57 NGLDGEPAVNADAAALHEDKAVYDQOLKKAGDNPYLELYNNADAFQORLJEDBTFSGNLGG 116
Db 12 NSLDGEPFNBSDDAAKHEDEAVDQYIKSGKNPYLEYFSPADQRFIDQTDADKMGKGVGH 71
QY 117 AVFOAKKRVLEBPLGVEBGAKTAPGKKRPEVEOSPOBDDSSGIGKTGOQ--PA----- 167
Db 72 YFFRTKRAFAFPLSTDBSE-----PG-----TSGVSRGKRTKPPAHIFVNO 112
QY 168 ----KKRLNFGOTGDSSEVPDPQPLGEPATPAVGPPTTMASSGGGAPMADNNEGALGVEN 223
Db 113 ARAKKRRASLAAQORLTLTWSDETETNQPTGTJANARVERSDGGS-SGGGSGGGGIGV 171
QY 224 ASGNHCOSTW--LDGRVITTSRTWAL-----PTYNNHLYKQJISASTGA 267
Db 172 STGTYDNOQTYTFYFLDGNVEITAHASRLHLHGMPSEBNTCRYVHNNQTTGHGTYKXGAM 231
QY 268 SNDNHVFGYSTWGYFDFNRFHCHSPSPDMORLINNMGFRFKLNFJLNIQVEKVT- 326
Db 232 AYDTHQQTW-TFMSLVDAWAMGWFPQSDWQFIONMESLNLDSLQELFNVVVKTYTBEQ 290
QY 327 ----NDGVTTLANNLTSTYQVFSDSYQLPYVLGSHQGL-----PPRPAD--VFMI 373
Db 291 OGAGODAKVYNNDLTACMMVALDSNNLLPYTPAAQTSBTLGEYPMKFPAPAPRYRYFFW 350
QY 374 POYGVLTLLNNGSO-----AVG-----RSSFYCLE-YFPSOMLRTGNNTFTF-SYTFEEVAP 420
Db 351 PROLSVTSNSNABEGQOITDTTIGEPQALNSQFFIENTLPLTLRTGDEPFTGTYIENITDP 410
QY 421 FHSSYA--*--HSOSLDRLMN-PLIDVLYLNTRTNOSGSAQNKDLF----- 463
Db 411 LKLTHWQNNRLACQGITDLPSTDTAVASLFLANGDRGSGTQTVNVNYTEALRTTRPAQ 470
QY 464 -----SRGSPAGMSVQPKWMLPQPCYRQORVSKTIDNNNSNFTTGASKYNL 511
Db 471 IGFMOPHDNFEANRGGPFKVPYVP-----LDITGEDHDAN---GAIRFVN 513
QY 512 NGR--ESTIINPCTA-----MASHDDDEKPEFPMGSGWMLFGKESAGASNTALDNYW 559
Db 514 GKQHGEDMAKQAPERYTWDAIDSAAGDYARCFV-----QSAPISTIPPNOQI- 563
QY 560 ITDEEERIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGWVQODRDVYLQGP1 618
Db 564 LQREBALAGRTMHTTNVNSGPELSAFPHPD-----IYNGOI 603
QY 619 WAK---IPHTDGHFHDPLMGGEGLKNPPOLLIKNTPYVPAWPAEF--SAYKFASTIO 673
Db 604 WDKELDLLEKPRPLHATAPFV---CKNNPFGQLFVHLGP---NLTDQFDPNSTIVSRIVT- 656
QY 674 YSTGOVSVEIEMELQKENSKNRNPQVQYNSN 704
Db 657 YSTFYWGKILKFKAKURPULTNWPNYQATID 687

```

Search completed: July 17, 2003, 18:32:02
Job time : 17.5704 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 ; Search time 54.7373 Seconds
(without alignments)
2770.518 Million cell updates/sec

Title: US-09-807-802A-13
Perfect score: 3989
Sequence: 1 MAADGTLPPWLEMDNLSGIR.....NNGLYTPRPICGRVYTRPL 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp Unclassified: *
15: sp_virus: *
16: sp_bacterioph: *
17: sp_archaeop: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3989	100.0	736	12	Q9WBP8 adeno-ssoc
2	3963	99.3	736	12	Q56137 adeno-ssoc
3	3511	88.0	736	12	Q56139 adeno-ssoc
4	3494	87.6	736	12	Q65311 adeno-ssoc
5	3402.5	85.3	735	12	Q56652 adeno-ssoc
6	2759.5	69.2	598	12	Q56653 adeno-ssoc
7	2486.5	62.3	734	12	Q41855 adeno-ssoc
8	2481.5	62.2	533	12	Q92917 adeno-ssoc
9	2220	55.7	724	12	Q9Y1J1 adeno-ssoc
10	2190.5	54.9	732	12	Q83290 muscovy duc
11	2173.5	54.5	732	12	Q67666 goose parvo
12	2168.5	54.4	732	12	Q83289 muscovy duc
13	2165.5	54.3	732	12	Q8V395 goose parvo
14	2162.5	54.2	732	12	Q65444 barbatie du
15	1990.5	49.9	676	12	Q67672 goose parvo
16	1712	42.9	587	12	Q67667 goose parvo

17	1698	42.6	534	12	Q67668 goose parvo
18	1695	42.5	587	12	Q65445 barbatie du
19	1676	42.0	534	12	Q65446 barbatie du
20	859.5	21.5	703	12	Q8QV5 minute vltu
21	715	17.9	179	12	Q9WA24 goose parvo
22	715	17.9	179	12	Q9W807 duck parvo
23	711	17.9	179	12	Q9WN18 duck parvo
24	709	17.8	179	12	Q9WN19 duck parvo
25	707	17.7	179	12	Q9WN20 goose parvo
26	618	15.5	571	12	Q8QV4 minute vltu
27	617	15.5	785	12	Q950X4 pig-tailed
28	614.5	15.4	781	12	Q9PZT0 hamster par
29	611	15.3	981	12	Q8QV11 hamster par
30	608.5	15.3	781	12	Q9QGP8 human eryth
31	606.5	15.2	781	12	Q912B8 human eryth
32	606.5	15.2	781	12	P89317 human parvo
33	606.5	15.2	781	12	P89318 human parvo
34	606.5	15.2	781	12	P89319 human parvo
35	604.5	15.2	773	12	Q913X1 human parvo
36	604.5	15.2	781	12	P89316 human parvo
37	603.5	15.1	781	12	P90221 human parvo
38	603.5	15.1	781	12	P90222 human parvo
39	603.5	15.1	781	12	P90223 human parvo
40	603.5	15.1	781	12	P90224 human parvo
41	603.5	15.1	781	12	Q85117 human parvo
42	603.5	15.1	781	12	Q90200 human parvo
43	603	15.1	829	12	P87584 chipmunk pa
44	602.5	15.1	781	12	Q91980 human parvo
45	601.5	15.1	781	12	P89321 human parvo

ALIGNMENTS

RESULT 1					
Q9WBP8	PRELIMINARY;	PRT;	736 AA.		
AC Q9WBP8;					
DT 01-NOV-1999 (TREMBLrel. 12, Created)					
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE Capsid protein.					
OC adeno-associated virus 1.					
OC Viruses; ssDNA viruses; Parvoviridae; Dependovirus.					
OX NCBI_TaxID=85106;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA MEDLINE=9214338; PubMed=10196295;					
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;					
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.					
DR EMBL: AF063497; AAD27757.1; -					
DR InterPro: IPR001403; Parvo_coat.					
DR Pfam: PF00740; Parvo_coat; 1.					
SQ SEQUENCE 736 AA; 81375 MW; CFAFB9BDSOD595 CRC64;					
Query Match	100.0%;	Score 3989;	DB 12;	Length 736;	
Best Local Similarity	100.0%;	Pred. No. 7.9e-258;			
Matches 736;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1 MAADGTLPPWLEMDNLSGIRWMDLPGAPKPKANQKDDGGLVLPGYKYLGPFGND 60					
DB 1 MAADGTLPPWLEMDNLSGIRWMDLPGAPKPKANQKDDGGLVLPGYKYLGPFGND 60					
QY 61 KGEPPVAAADAAALHKKAYDQQLKAGDNYLRYNADAFQERLQDTSFGGLGAAVQ 120					
DB 61 KGEPPVAAADAAALHKKAYDQQLKAGDNYLRYNADAFQERLQDTSFGGLGAAVQ 120					
QY 121 AKKRVLEPLGLVEBAKTAPEGKRPVEQSPOEPDSSSGIGTKGQCPAKRLNFGCGDSE 180					

```

Db 121 AKKRVLEPLGLVEEGAKTAPKRRPVEQSPQEPDSSSGIGKGTQQAKKLNFQGTGDSSE 180
Qy 181 SVDPDPLGEPPTPAVGGPTTMAAGGAPMADNNEGADGVANASGWHCDSTWLGDRVI 240
Db 181 SVDPDPLGEPPTPAVGGPTTMAAGGAPMADNNEGADGVANASGWHCDSTWLGDRVI 240
Qy 241 TTSTRMALPTYNHLYKQISSASTGASNDNHFGYSTPWGTFDFNRFCHFSRDMQRL 300
Db 241 TTSTRMALPTYNHLYKQISSASTGASNDNHFGYSTPWGTFDFNRFCHFSRDMQRL 300
Qy 301 INNMGFRPRKLNFKLNIOVKEVTTNDGVTIANNLSTVOVFSDEYOLPYVLSAQH 360
Db 301 INNMGFRPRKLNFKLNIOVKEVTTNDGVTIANNLSTVOVFSDEYOLPYVLSAQH 360
Qy 361 GCLPPPADVFMIPQGYLTLLNNGSOAVGRSSFCLEYPFSDMLRTGNFTFSYTFEEVP 420
Db 361 GCLPPPADVFMIPQGYLTLLNNGSOAVGRSSFCLEYPFSDMLRTGNFTFSYTFEEVP 420
Qy 421 FHSSVYASQSLDRMLNPLIDQYLYLNRTONGSQAQNDLFSRGSAGMSVQPKMWP 480
Db 421 FHSSVYASQSLDRMLNPLIDQYLYLNRTONGSQAQNDLFSRGSAGMSVQPKMWP 480
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMG 540
Db 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMG 540
Qy 541 MIFKESAGASNTALDNVMTTDEBEIKATNPVATERFGTYAVNFQSSSTDPATGVAMG 600
Db 541 MIFKESAGASNTALDNVMTTDEBEIKATNPVATERFGTYAVNFQSSSTDPATGVAMG 600
Qy 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFSPMLMGFGGLKNDPOLLKNTFVPANPPA 660
Db 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFSPMLMGFGGLKNDPOLLKNTFVPANPPA 660
Qy 661 EFSAITKFASTITQYSTGVSEIEMELQKENSKRANPEVOYTSNYAKSANVDFTVNNGL 720
Db 661 EFSAITKFASTITQYSTGVSEIEMELQKENSKRANPEVOYTSNYAKSANVDFTVNNGL 720
Qy 721 YTEPRPIGTRYLRPL 736
Db 721 YTEPRPIGTRYLRPL 736

```

RESULT 2

OS6137 PRELIMINARY; PRT; 736 AA.

AC 056137; TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE Capsid protein VP1.

OS adeno-associated virus 6.

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.

OX NCBI_TaxID=68558;

RN [1]

RP SEQUENCE FROM N.A.

RA Rutledge E.A., Russel D.W.,

RT "Infectious clones and vectors derived from adeno-associated virus

RT (AAV) serotypes other than AAV type 2."

RL J. Virol. 72:309-319 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RA Rutledge E.A., Russel D.W.,

RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF028704; AAB95450.1;

DR InterPro; IPR001403; Parvo coat.

DR Pfam; PF00740; Parvo coat.1.

SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

Query Match 99.3%; Score 3963; DB 12; Length 736;

Best Local Similarity 99.2%; Pred. No. 4.3e-256;

```

Matches 730; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MAADGYLPWLENNLESEGRREWWDLKPGAPKPAKNOQKODGGLVPGKYTGPPENGLD 60
Db 1 MAADGYLPWLENNLESEGRREWWDLKPGAPKPAKNOQKODGGLVPGKYTGPPENGLD 60
Qy 61 KGPVNAADAAALEHKAAYDQOLKAGDNPLYRNHADAEOERLOEDTSFGNLGRAVFO 120
Db 61 KGPVNAADAAALEHKAAYDQOLKAGDNPLYRNHADAEOERLOEDTSFGNLGRAVFO 120
Qy 121 AKKRVLEPLGLVEEGAKTAPKRRPVEQSPQEPDSSSGIGKGTQQAKKLNFQGTGDSSE 180
Db 121 AKKRVLEPLGLVEEGAKTAPKRRPVEQSPQEPDSSSGIGKGTQQAKKLNFQGTGDSSE 180
Qy 181 SVDPDPLGEPPTPAVGGPTTMAAGGAPMADNNEGADGVANASGWHCDSTWLGDRVI 240
Db 181 SVDPDPLGEPPTPAVGGPTTMAAGGAPMADNNEGADGVANASGWHCDSTWLGDRVI 240
Qy 241 TTSTRMALPTYNHLYKQISSASTGASNDNHFGYSTPWGTFDFNRFCHFSRDMQRL 300
Db 241 TTSTRMALPTYNHLYKQISSASTGASNDNHFGYSTPWGTFDFNRFCHFSRDMQRL 300
Qy 301 INNMGFRPRKLNFKLNIOVKEVTTNDGVTIANNLSTVOVFSDEYOLPYVLSAQH 360
Db 301 INNMGFRPRKLNFKLNIOVKEVTTNDGVTIANNLSTVOVFSDEYOLPYVLSAQH 360
Qy 361 GCLPPPADVFMIPQGYLTLLNNGSOAVGRSSFCLEYPFSDMLRTGNFTFSYTFEEVP 420
Db 361 GCLPPPADVFMIPQGYLTLLNNGSOAVGRSSFCLEYPFSDMLRTGNFTFSYTFEEVP 420
Qy 421 FHSSVYASQSLDRMLNPLIDQYLYLNRTONGSQAQNDLFSRGSAGMSVQPKMWP 480
Db 421 FHSSVYASQSLDRMLNPLIDQYLYLNRTONGSQAQNDLFSRGSAGMSVQPKMWP 480
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMG 540
Db 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMG 540
Qy 541 MIFKESAGASNTALDNVMTTDEBEIKATNPVATERFGTYAVNFQSSSTDPATGVAMG 600
Db 541 MIFKESAGASNTALDNVMTTDEBEIKATNPVATERFGTYAVNFQSSSTDPATGVAMG 600
Qy 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFSPMLMGFGGLKNDPOLLKNTFVPANPPA 660
Db 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFSPMLMGFGGLKNDPOLLKNTFVPANPPA 660
Qy 661 EFSAITKFASTITQYSTGVSEIEMELQKENSKRANPEVOYTSNYAKSANVDFTVNNGL 720
Db 661 EFSAITKFASTITQYSTGVSEIEMELQKENSKRANPEVOYTSNYAKSANVDFTVNNGL 720
Qy 721 YTEPRPIGTRYLRPL 736
Db 721 YTEPRPIGTRYLRPL 736

```

RESULT 3

OS6139 PRELIMINARY; PRT; 736 AA.

AC 056139; TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE Capsid protein VP1.

OS adeno-associated virus 3B.

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.

OX NCBI_TaxID=68742;

RN [1]

RP SEQUENCE FROM N.A.

RA Rutledge E.A., Russel D.W.,

RT "Infectious clones and vectors derived from adeno-associated virus

RT (AAV) serotypes other than AAV type 2."

RL J. Virol. 72:309-319 (1998).

[2]
 RN SEQUENCE FROM N.A.
 RP Rutledge E.A., Russell D.W.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF028705; AAB95452.1; -
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;

Query Match 88.0%; Score 3511; DB 12; Length 736;
 Best Local Similarity 86.8%; Pred. No. 6.2e-226;
 Matches 640; Conservative 38; Mismatches 57; Indels 2; Gaps 2;

```

QY 1 MAADGTLPLWLENLSEGIKREMDLKPAPKPKANQKODDGRGLVPGYKYLGPFGD 60
DB 1 MAADGTLPLWLENLSEGIKREMDLKPAPKPKANQKODDGRGLVPGYKYLGPFGD 60
QY 61 KGEPVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEFOELQEDTSFGNLGRAVFO 120
DB 61 KGEPVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEFOELQEDTSFGNLGRAVFO 120
QY 121 AKKRVLPELGLVEEAKTAPGKKRPVQSPQEPDSSSGIKTQOQAPKRLNFGOTDSE 180
DB 121 AKKRVLPELGLVEEAKTAPGKKRPVQSPQEPDSSSGIKTQOQAPKRLNFGOTDSE 180
QY 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNEGADGVNAGSNWCHDSQWLDGRVI 240
DB 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNEGADGVNAGSNWCHDSQWLDGRVI 240
QY 241 TTSRTMALPTYNHLYKOISSASTGASNDNHFGYSTPWGYDFNRFCHFSRDMORL 300
DB 241 TTSRTMALPTYNHLYKOISSASTGASNDNHFGYSTPWGYDFNRFCHFSRDMORL 300
QY 301 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
DB 301 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
QY 361 GCLPPPADVFMIPQYGYLTNNNSQAVGRSSFYCLEYFSPQMLRGNNTFTSYTEEDVP 420
DB 361 GCLPPPADVFMIPQYGYLTNNNSQAVGRSSFYCLEYFSPQMLRGNNTFTSYTEEDVP 420
QY 421 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
DB 421 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
QY 479 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
DB 479 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
QY 480 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
DB 480 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
QY 539 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
DB 539 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
QY 540 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
DB 540 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
QY 599 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
DB 599 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
QY 600 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
DB 600 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
QY 659 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
DB 659 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
QY 660 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
DB 660 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
QY 719 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
DB 719 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
QY 720 LYTEPRPPIGRYITRRL 736
DB 720 LYTEPRPPIGRYITRRL 736
QY 736 LYTEPRPPIGRYITRRL 736
DB 736 LYTEPRPPIGRYITRRL 736

```

RESULT 4
 065311 PRELIMINARY; PRT; 736 AA.
 AC 065311;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)

DE Capsid protein.
 OS adeno-associated virus 3.
 OC viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=46350;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H;
 RX MEDLINE=96266430; PubMed=8661429;
 RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
 RT "Nucleotide sequencing and generation of an infectious clone of adeno-
 associated virus 3."
 RL Virology 221:208-217(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H;
 RA Muramatsu S., Brown K.E.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U48704; AAC55049.1; -
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 736 AA; 81660 MW; AF1EPA7B5C67A10 CRC64;

Query Match 87.6%; Score 3494; DB 12; Length 736;
 Best Local Similarity 86.4%; Pred. No. 8.5e-225;
 Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;

```

QY 1 MAADGTLPLWLENLSEGIKREMDLKPAPKPKANQKODDGRGLVPGYKYLGPFGD 60
DB 1 MAADGTLPLWLENLSEGIKREMDLKPAPKPKANQKODDGRGLVPGYKYLGPFGD 60
QY 61 KGEPVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEFOELQEDTSFGNLGRAVFO 120
DB 61 KGEPVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEFOELQEDTSFGNLGRAVFO 120
QY 121 AKKRVLPELGLVEEAKTAPGKKRPVQSPQEPDSSSGIKTQOQAPKRLNFGOTDSE 180
DB 121 AKKRVLPELGLVEEAKTAPGKKRPVQSPQEPDSSSGIKTQOQAPKRLNFGOTDSE 180
QY 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNEGADGVNAGSNWCHDSQWLDGRVI 240
DB 181 SVDPDPLGEPPTPAVGPPTMAAGGAPMADNNEGADGVNAGSNWCHDSQWLDGRVI 240
QY 241 TTSRTMALPTYNHLYKOISSASTGASNDNHFGYSTPWGYDFNRFCHFSRDMORL 300
DB 241 TTSRTMALPTYNHLYKOISSASTGASNDNHFGYSTPWGYDFNRFCHFSRDMORL 300
QY 301 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
DB 301 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
QY 360 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
DB 360 INNWGFRPKLNFKLFNIQVEKVTNDGVTIANNLSTVOYFSDSEYOLPYLGSAGQ 360
QY 421 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
DB 421 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
QY 479 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
DB 479 FHSVYASOSLDLMLPLIDQYLYLNRTQ-NOSGSAOKKDLFSRGSPPAGMSVOPKRWL 479
QY 480 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
DB 480 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
QY 539 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
DB 539 PGPCYRQORLSKTKANDNNNSNFTWTGASKYLNLRGSIINPGTAMASHKODEKFFPMHG 539
QY 540 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
DB 540 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
QY 599 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
DB 599 VMIFGESAGASTALDNNMTDEEIKATNPVATERFGTVAVNFOSSSTDPAIDGVHAM 599
QY 600 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
DB 600 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
QY 659 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
DB 659 GALPGWMDRDVYLGGPIWAKIPHTDGHFSPPLMGFGGLKPNPQIILIKNTVPANP 659
QY 660 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
DB 660 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
QY 719 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719
DB 719 AEFSAATKASFTIOYSTGVSVIEIEMELQKENSKRANPEVOYTSNYSKSNVDFVTDNG 719

```

QY 720 LYBPRPIGRYLRPL 736
Db 720 YSBPRPIGRYLRNL 736

RESULT 5

056652 PRELIMINARY; PRT; 735 AA.
AC 056652;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Major coat protein VP1.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
ON NCBI_TaxID=10804;
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF043303; AAC03780.1;
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KM Coat protein.
FT VARIANT 76 76 D -> V.
FT VARIANT 553 553 D -> N.
FT VARIANT 567 567 T -> N.
FT VARIANT 677 678 QV -> HV.
FT VARIANT 710 710 V -> R.
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match Best Local Similarity 85.3%; Score 3402.5; DB 12; Length 735;
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDWLEEDNLSSGIREWDLKPGAPKPKANQKOD3RGVLVPGYKYLCPENGGLD 60
Db 1 MAADGYLPDWLEEDNLSSGIREWDLKPGAPKPKANQKOD3RGVLVPGYKYLCPENGGLD 60
QY 61 KGEVYNADAAALHDKAYDQQLKAGNPNRYRNHADAEPQERLOEPTSGGNGRAVFO 120
Db 61 KGEVYNADAAALHDKAYDQQLKAGNPNRYRNHADAEPQERLOEPTSGGNGRAVFO 120
QY 121 AKKRVLEPLGLVEBGAATAPGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDB 180
Db 121 AKKRVLEPLGLVEBGAATAPGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDB 180
QY 121 AKKRVLEPLGLVEBGAATAPGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDB 180
Db 121 AKKRVLEPLGLVEBGAATAPGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDB 180
QY 181 SVDPDQPLGEPAPPAVGPPTMASSGGAPADNNEGADGVGNASGNWCHDSTLGRVY 240
Db 181 SVDPDQPLGEPAPPAVGPPTMASSGGAPADNNEGADGVGNASGNWCHDSTLGRVY 240
QY 181 SVDPDQPLGEPAPPAVGPPTMASSGGAPADNNEGADGVGNASGNWCHDSTLGRVY 240
Db 181 SVDPDQPLGEPAPPAVGPPTMASSGGAPADNNEGADGVGNASGNWCHDSTLGRVY 240
QY 241 TTSRTMALPTNNHLYKQISSASTGASNDNHYFGYSTPMGTFPNNFCHFSRDMOQL 300
Db 241 TTSRTMALPTNNHLYKQISSASTGASNDNHYFGYSTPMGTFPNNFCHFSRDMOQL 300
QY 241 TTSRTMALPTNNHLYKQISSASTGASNDNHYFGYSTPMGTFPNNFCHFSRDMOQL 300
Db 241 TTSRTMALPTNNHLYKQISSASTGASNDNHYFGYSTPMGTFPNNFCHFSRDMOQL 300
QY 301 INNMGGRPKRLNFKLNIQYKEVTTNDGVTIANNTLSTVQVSDSEYQLPYVLSAHQ 360
Db 301 INNMGGRPKRLNFKLNIQYKEVTTNDGVTIANNTLSTVQVSDSEYQLPYVLSAHQ 360
QY 300 INNMGGRPKRLNFKLNIQYKEVTTNDGVTIANNTLSTVQVSDSEYQLPYVLSAHQ 359
Db 300 INNMGGRPKRLNFKLNIQYKEVTTNDGVTIANNTLSTVQVSDSEYQLPYVLSAHQ 359
QY 361 GCLPPPADVFMATQYGLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTSTFEEDVP 420
Db 361 GCLPPPADVFMATQYGLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTSTFEEDVP 419

QY 421 FHSYAHSSGLDRLMPLIDQYLYINRTONOSGASQNKDLPSRGSAGASVQPKWLP 480
Db 420 FHSYAHSSGLDRLMPLIDQYLYINRTONOSGASQNKDLPSRGSAGASVQPKWLP 479
QY 481 GPCYRQORVSKRTDNNNSNFTWTGASKYNLANGRESIINGTAMASHKDEKFFPMG 540
Db 480 GPCYRQORVSKRTDNNNSNFTWTGASKYNLANGRESIINGTAMASHKDEKFFPMG 539
QY 541 MITGESAGASNLADNMTTDEEIKATNPVATERFGYAVVFOSSSTDPACDVHMG 600
Db 540 LIFKQGSSEKTNDIEKVMITDEEIRTNVPATEGYGVSTYLQRGNRAADAVDTQG 599
QY 601 ALPGMWQDQDVYLOQPIYAKIPIHTDGHFSPPLMGFGKPNPOLLKNTVPANPPA 660
Db 600 VLFQMWQDQDVYLOQPIYAKIPIHTDGHFSPPLMGFGKPNPOLLKNTVPANPPA 659
QY 661 EFSATKPSFTTQSTGOVSEIEWELQKNSKRNMPDEVQYTSNVAKANVDETVNGL 720
Db 660 TFSAAKPSFTTQSTGOVSEIEWELQKNSKRNMPDEVQYTSNVAKANVDETVNGL 719
QY 721 YTBPRPIGRYLRPL 736
Db 720 YSBPRPIGRYLRNL 736

RESULT 6

056653 PRELIMINARY; PRT; 598 AA.
AC 056653;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Major coat protein VP2.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
ON NCBI_TaxID=10804;
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF043303; AAC03778.1;
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KM Coat protein.
FT VARIANT 416 416 D -> N.
FT VARIANT 430 430 T -> N.
FT VARIANT 540 541 QV -> HV.
FT VARIANT 573 573 V -> R.
SQ SEQUENCE 598 AA; 66619 MW; 070811ED9368B34 CRC64;

Query Match Best Local Similarity 69.2%; Score 2759.5; DB 12; Length 598;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDBSEVPDPQIGEPAPPAV 198
Db 2 APGKKRPVEQSPOEPDSSSGIGTGOQAPAKRLNFGQTSDBSEVPDPQIGEPAPPAV 198
QY 199 GPTTMASSGAPADNNEGADGVGNASGNWCHDSTLGRVITTSRTMALPTNNHLYK 258
Db 62 GPTTMASSGAPADNNEGADGVGNASGNWCHDSTLGRVITTSRTMALPTNNHLYK 121

QY 259 QISSASTGASNDNHFGYSTPWGTFEDNRRHCHFSPPDMORLNNNNMFRPKLNFLEN 318
 Db 122 QISSQS -GASNDNHFGYSTPWGTFEDNRRHCHFSPPDMORLNNNNMFRPKLNFLEN 180
 QY 319 IQYKVTNDGVTIANNLSTVOVFSDSYOLPYVLSAHOGCLPPPADVEMIPQYGY 378
 Db 181 IQYKVTNDGVTIANNLSTVOVFSDSYOLPYVLSAHOGCLPPPADVEMIPQYGY 240
 QY 379 LTLNNGSOAVGRSSFCLEYFPPSOMLTGNNFFSYFEEVPHSSVAHSOSLDRLNPL 438
 Db 241 LTLNNGSOAVGRSSFCLEYFPPSOMLTGNNFFSYFEEVPHSSVAHSOSLDRLNPL 300
 QY 439 IDOYLVLNTONOSGSAQNKDILLFSRGSFAGMSVQPKNMLPGFCYRQOQVSKITDNNN 498
 Db 301 IDOYLVLNTONOSGSAQNKDILLFSRGSFAGMSVQPKNMLPGFCYRQOQVSKITDNNN 360
 QY 499 SNFTWGTASRYNLNGRESIINPGTAMASHDDEDEKFFPMGSGVMI FKGESAGASNTALDNV 558
 Db 361 SEVSWTGTATKYLHNGRSLVNPFGPAMASHKDEDEKFFPMGSGVMI FKGESAGASNTALDNV 420
 QY 559 MINDDEEIKATNPATERFGTAVANFPOSSSTDPAATGVDHMGALPGWVMDRDTVYLCGP 618
 Db 421 MINDDEEIKATNPATERFGTAVANFPOSSSTDPAATGVDHMGALPGWVMDRDTVYLCGP 480
 QY 619 WAKIPHTDGHFHSPLMGFGGLKNPPQIILKTPVPANPFAESATKPFASFTIYSTGQ 678
 Db 481 WAKIPHTDGHFHSPLMGFGGLKNPPQIILKTPVPANPFAESATKPFASFTIYSTGQ 540
 QY 679 VSVEIEMELOKENSKEKNPEVOYTSNYAKSANDFTVNDNGLYTEPRPIGTIRYLTRPL 736
 Db 541 VSVEIEMELOKENSKEKNPEVOYTSNYAKSANDFTVNDNGLYTEPRPIGTIRYLTRPL 598

RESULT 7

QY 041855 PRELIMINARY; PRT; 734 AA.
 AC 041855;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE Capsid.
 OS adeno-associated virus 4.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 NC NCB1_TaxID=57579;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC VR-646;
 RA MEDLINE=97404695; PubMed=9261407;
 RX Chlorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;
 RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of
 RT recombinant AAV4 particles";
 RL J. Virol. 71:6823-6833(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC VR-646;
 RA Chlorini J.A., Yang L., Kotin R.M., Safer B.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U89790; AAC8045.1; -;
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat.1.
 SQ SEQUENCE 734 AA; 80639 MW; 616CC27A77BBEGF CRC64;

Query Match 62.3%; Score 2486.5; DB 12; Length 734;
 Best Local Similarity 63.5%; Pred. No. 1,4e-157;
 Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

QY 4 DGYLPDMLIEDNLSEGLREWMDLKGAEPKPRANOQKQDQGLVLPYKYLGPFNGLDKGE 63
 Db 3 DGYLPDMLIEDNLSEGLREWMALDQGAEPKPRANOQHONAGVLPYKYLGPFNGLDKGE 62
 QY 64 PVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAFOERLQEDTSGFNGILGRAVQAKK 123
 Db 63 PVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAFOERLQEDTSGFNGILGRAVQAKK 122

QY 124 RVLEPLGLVEEAGAKTAPGKKRPVEQSPDSSSGIGTGOQPAKKRLNF-GOTGDSSEV 182
 Db 123 RVLEPLGLVEEAGAKTAPGKKRPVEQSPDSSSGIGTGOQPAKKRLNFEDTGTAGDGP 182
 QY 183 PDBPGLGEPPEPATPAVAGPTTASGGAPADNNEGADGVGNAGSNHCHDSYTWLGDVYTT 242
 Db 183 PEGSTSG-----AMSDSEMRRAAGAAVBEQGQADGVGNAGSNHCHDSYTWLGDVYTT 236
 QY 243 STRTMALPTYNMLLYQISSASTGASNDNHFGYSTPWGTFEDNRRHCHFSPPDMORLIN 302
 Db 237 STRTMALPTYNMLLYQISSASTGASNDNHFGYSTPWGTFEDNRRHCHFSPPDMORLIN 292
 QY 303 NNGFPRPKLNFLENIQYKVTNDGVTIANNLSTVOVFSDSYOLPYVLSAHOGCL 362
 Db 293 NNGFPRPKLNFLENIQYKVTNDGVTIANNLSTVOVFSDSYOLPYVLSAHOGCL 352
 QY 363 LPPPADVEMIPQYGY---LTLNNGSOAVGRSSFCLEYFPPSOMLTGNNFFSYFEEV 419
 Db 353 LPPPADVEMIPQYGY---LTLNNGSOAVGRSSFCLEYFPPSOMLTGNNFFSYFEEV 412
 QY 420 PFHSSVAHSOSLDRLNPLIDQYLTYLNTON---QSGSAQNKDILLFSRGSFAGMSVQ 475
 Db 413 PFHSSVAHSOSLDRLNPLIDQYLTYLNTON---QSGSAQNKDILLFSRGSFAGMSVQ 469
 QY 476 KMWLPDPCYRQOQVSKITDNNNFTWTGAS---RY---NNGRESIINPGTAMASHK 528
 Db 470 KMWLPDPCYRQOQVSKITDNNNFTWTGAS---RY---NNGRESIINPGTAMASHK 527
 QY 529 DDEDEKFFPMGSGVMI FKGESAGASNTALDNVMTDEEIKATNPATERFGTAVANFPOSS 588
 Db 528 DDEDEKFFPMGSGVMI FKGESAGASNTALDNVMTDEEIKATNPATERFGTAVANFPOSS 586
 QY 589 TDPATGVDHMGALPGWVMDRDTVYLCGPWAKIPHTDGHFHSPLMGFGGLKNPPQIIL 648
 Db 587 TDPATGVDHMGALPGWVMDRDTVYLCGPWAKIPHTDGHFHSPLMGFGGLKNPPQIIL 646
 QY 649 IKATPVANPFAESATKPFASFTIYSTGQVSVEIEMELOKENSKEKNPEVOYTSNYAKS 708
 Db 647 IKATPVANPFAESATKPFASFTIYSTGQVSVEIEMELOKENSKEKNPEVOYTSNYAKS 706
 QY 709 ANVDFTVNDNGLYTEPRPIGTIRYLTRPL 736
 Db 707 NSLLMAPDAAGKTEBRATIGTRYLTHL 734

RESULT 8

QY 092917 PRELIMINARY; PRT; 533 AA.
 AC 092917;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE Major coat protein VP3.
 OS Adeno-associated virus 2 (AAV2).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 NC NCB1_TaxID=10804;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95088582; PubMed=7996133;
 RA Ruffing M., Heid H., Kleinschmidt J.A.;
 RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
 RT proteins affect viral infectivity: lack of an RGD integrin-binding
 RT motif";
 RL J. Gen. Virol. 75:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Berne K.I., Bohenkly R.A., Cassinotti P., Colvin D., Donahue B.A.,
 RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
 RL Tratschin J.-D., Weitz M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF043303; AAC03779.1; -;
 DR InterPro: IPR001403; Parvo_coat.

DR Pfam; PF00740; Parvo_coat; 1.
 KM Cost Protein 351 351 D -> N.
 FT VARIANT 365 365 T -> N.
 FT VARIANT 475 476 QV -> HV.
 FT VARIANT 508 508 V -> R.
 SQ SEQUENCE 533 AA; 60063 MW; 9E4DB8C25810D4F0 CRC64;

Query Match 62.2%; Score 2481.5; DB 12; Length 533;
 Best Local Similarity 83.3%; Pred. No. 1.9e-157;
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 203 MASGGAAPMADNNGADGVNAGSNWHDSTWMLGDRVITSTRTWALPTYNHLYKOISS 262
 DB 1 MATSGAPMADNNGADGVNAGSNWHDSTWMLGDRVITSTRTWALPTYNHLYKOISS 60
 QY 263 ASTGASDNHFGYSTPWGYDFNRFCHFSPPDMQRLINNNGFRPKLNFNIQVK 322
 DB 61 QS-GASNDNHFGYSTPWGYDFNRFCHFSPPDMQRLINNNGFRPKLNFNIQVK 119
 QY 323 EYTNNDGYTTIANNLSTVOYFSPSEYOLPYVLGSAHQCLFPPADVFMTPQYGYLTN 382
 DB 120 EYTNNDGYTTIANNLSTVOYFSPSEYOLPYVLGSAHQCLFPPADVFMTPQYGYLTN 179
 QY 383 NSGAVGRSSPYCLEYFSPQMLRTGNNFTFSYFEEVFPFSSYAHSGSLDLAMPIDQY 442
 DB 180 NSGAVGRSSPYCLEYFSPQMLRTGNNFTFSYFEEVFPFSSYAHSGSLDLAMPIDQY 239
 QY 443 LYLINRTQNGSAGNKKDLFRSGSPAGMSVOPKMWLPGPCYRQORVSKTKTDNNNSFT 502
 DB 240 LYLINRTQNGSAGNKKDLFRSGSPAGMSVOPKMWLPGPCYRQORVSKTKTDNNNSFT 299
 QY 503 WTGASKYLNRESITINPGTAMASHKDEDEKFPMSGIMTCKESAGSNALNNMTD 562
 DB 300 WTGASKYLNRESITINPGTAMASHKDEDEKFPMSGIMTCKESAGSNALNNMTD 359
 QY 563 EEEKATNPVATEFPGYAVNFGSSSTDPATGVNAGALPCMWODRDVYLQCPFWAKI 622
 DB 360 EEEKATNPVATEFPGYAVNFGSSSTDPATGVNAGALPCMWODRDVYLQCPFWAKI 419
 QY 623 PHTDGHFSPPLMGFGKLPNDPOLLKNTTPVANPAPASATYRASFITQYSTGQSV 682
 DB 420 PHTDGHFSPPLMGFGKLPNDPOLLKNTTPVANPAPASATYRASFITQYSTGQSV 479
 QY 683 IEMELOKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTEPRPIGRYLRPL 736
 DB 480 IEMELOKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTEPRPIGRYLRPL 533

RESULT 9

QY 09YJ1 PRELIMINARY; PRT; 724 AA.
 AC 09YJ1;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Capsid protein.
 GN CAP OR VP1.
 OS adeno-associated virus 5.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 NCBI TaxID=82300;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99098980; PubMed=9882294;
 RA Bantel-Schaal U., Delius H., Schmidt R., Zur Hausen H.,
 RT "Human adeno-associated virus type 5 is only distantly related to
 RT other known primate helper-dependent parvoviruses.";
 RL J. Virol. 73:939-947 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99099022; PubMed=9882336;
 RA Chiorini J.A., Kim F., Yang L., Kotin R.M.,
 RT "Cloning and characterization of adeno-associated virus type 5.";
 RT

RL J. Virol. 73:1309-1319 (1999);
 DR EMBL; Y18065; CA477024.1; -;
 DR EMBL; AF085716; AAD13756.1; -;
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 SQ SEQUENCE 724 AA; 80424 MW; B8C332B88258D34C CRC64;

Query Match 55.7%; Score 2220; DB 12; Length 724;
 Best Local Similarity 58.6%; Pred. No. 8.4e-140;
 Matches 431; Conservative 81; Mismatches 198; Indels 26; Gaps 9;

QY 8 PDWLEDNLSGIREWMDLKPAKPKXANQKODDGRGLVPGYKYGPFNGLDKGEVNA 67
 DB 8 PDWLEB-VEGRLREFGLBAGPKPKPNQHQOARGLVPGYNYLCPGNGLDGGEVNR 66
 QY 68 ADAALAEHDKAYOOLKAGNPFILRYNHADEFOERLQETSTSGNUGRAVFOAKRYLE 127
 DB 67 ADEVARHEDHISYNEOLEAGDNPLYKYNHADEFOERLADDTSEGNLGAFAKRVLE 126
 QY 128 PLGIVERGAKTAPGKKRPVQSPQEPDSSGICKTGQOPAKKRLNFQGTDSSEVPDP 187
 DB 127 PLGIVERGAKTAPGKKRIDHFKRKKART-----EEDSKP-----STSDAEAGBSG 176
 QY 188 LGEPATPA-AVGPPTMASGGAAPMADNNGADGVNAGSNWHDSTWMLGDRVITSTRT 246
 DB 177 QLOIPAPASSLADMTMSAGGGPLGDNNGADGVNAGSNWHDSTWMLGDRVITSTRT 236
 QY 247 WALPTYNHLYKOISSASTASANDNHFGYSTPWGYDFNRFCHFSPPDMQRLINNNG 306
 DB 237 WALPTYNHLYKOISSASTASANDNHFGYSTPWGYDFNRFCHFSPPDMQRLINNNG 296
 QY 307 FRPRLNFKLFNIQVEVTTNDGYTTIANNLSTVOYFSPSEYOLPYVLGSAHQCLPF 366
 DB 297 FRPRLNFKLFNIQVEVTTNDGYTTIANNLSTVOYFSPSEYOLPYVLGSAHQCLPF 356
 QY 367 PADVFMTPQYGYLTN--NSGAVGRSSPYCLEYFSPQMLRTGNNFTFSYFEEVFPFSS 424
 DB 357 PPOVFTLPQGYATVLRNDNTENTPERSFPCLYFSPSKMLRTGNNFTFSYFEEVFPFSS 416
 QY 425 YAHSGSLDLAMPIDQYLYLNRTQNGSAGNKKDLFRSGSPAGMSVOP-KMWLPGPC 483
 DB 417 FAPSNLFKLANPLVQYLYRFSYTNNTGVOFNKUL-----AGRYANTYNNKPPGM 469
 QY 484 YRQORVSKTKTDNNNSFTWTGASKYLNRESITINPGTAMASHKDEDEKFPMSGMIF 543
 DB 470 GRTQGMUGSGVBARASAFATNRMELBEGASQVPPQPMGNNNGSGSTYALEMTIF 529
 QY 544 GKESAGASNTAL--DNVMTDEEIKATNPVATERFGYAVNFGSSSTDPATGVNAG 600
 DB 530 NSQPANFGTATVLEGMLITSESETPVNRVAVNNGQWATNNGSTAPATGYVLOE 589
 QY 601 ALFGMWQODRDVYLQCPFWAKIHTDGHFSPPLMGFGKLPNDPOLLKNTTPVANP 660
 DB 590 IVPGSWMERDVLQGIYAKIPETGAHFSPSPAMGFGKLPNDPOLLKNTTPVANP 648
 QY 661 EFSATKPSATITQYSTGQSVIEIEMELOKENSKRANPEVOYTSNYAKSANVDFTV 720
 DB 649 SPSDVPASSITQYSTGQSVIEIEMELOKENSKRANPEVOYTSNYAKSANVDFTV 708
 QY 721 YTEPRPIGRYLRPL 736
 DB 709 YTRTRPIGRYLRPL 724

RESULT 10

QY 083290 PRELIMINARY; PRT; 732 AA.
 AC 083290;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Capsid protein.
 OS Muscovy duck parvovirus.

OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 RN NCBI_Taxid=37325;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=89384/France;
 RX MEDLINE=96406928; PubMed=8811015;
 RA Le Gall-Reculle G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
 RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)
 in a baculovirus expression system and demonstration of immunity
 induced by the recombinant protein.";
 RL J. Gen. Virol. 77:2159-2163 (1996).
 DR EMBL; Z68272; CA92575.1; -
 DR Interpro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo coat; 1.
 FT CHAIN 146 732 VP2 CAPSID PROTEIN.
 FT CHAIN 199 732 VP3 CAPSID PROTEIN.
 SQ SEQUENCE 732 AA; 81364 MW; D670DCCAB215F462 CRC64;

Query Match 54.9%; Score 2190.5; DB 12; Length 732;
 Best Local Similarity 55.7%; Pred. No. 7.9e-138;
 Matches 422; Conservative 88; Mismatches 184; Indels 63; Gaps 13;

9 DMLNDLSGIRREWMWLPKGPAPKANKOQK-----DDGRGLVLPKYYLGPEN 57
 10 DWYE-----TAAAGWRHLKAGAPKPKSNQOSQSVSTRKPKORKNNGFVLPGYKYLGPEN 65
 58 GLDKGPRVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEPORLEOEDTSFGNLGRA 117
 66 GLDKGPRVNAADVALEHDKAYDQOLKAGDNPYIKFNHADOEDTIDNLSQDTSFGNLGRA 125
 118 VFOAKRRVLEPLGLVEGAKTAPGKRRPVQSGFOEPDSSGIGKTGQOPAKKRLNFQQTG 177
 126 VFOAKRRVLEPLGLVEGAKTAPGKRRPVQSGFOEPDSSGIGKTGQOPAKKRLNFQQTG 177
 178 DSESVDPDPLG-----EPPTPAVGPPTMASGGGAPMADNNEGADGVGNASG 226
 171 --ENSSPSPSGGSAATAEGSEPVAA-----NMAEGSGAGMGSSAGAGAGVGNASG 222
 227 NMHCDSWTGADRYITTTSTRTWALPTYNHLYKQISSASTGASNDNHYFGYSTPWGFDEN 286
 223 NMHCDSQWGLDVTITKTRTWLPSYNNHLYKAITSGTNDPSN--TQYAGSTPWGFDEN 281
 287 RFHCSPSPDWORLNNNGFPRKRLNFKLNIQKVEVTNDGVTIANNLSTVOVPSD 346
 282 RFHCSPSPDWORLNNNGFPRKRLNFKLNIQKVEVTNDGVTIANNLSTVOVPSD 341
 347 SEYOLPYVIGSAGHOGCLPPFPADVEMIPQYGYLTN---NGSQAVERSSFYCLEYPPSQM 403
 342 NEHQLPYVIGSAGHOGCLPPFPADVEMIPQYGYLTN---NGSQAVERSSFYCLEYPPSQM 401
 404 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 463
 402 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 455
 402 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 455
 464 SRGSPAGMSVQPKNMLPGPCYRQORVSKTK--TDNNNSFTWTGASKYNLNGESLIINGP 521
 456 KKAIVKAGAFGAMGNMLPGPCYRQORVSKTK--TDNNNSFTWTGASKYNLNGESLIINGP 515
 522 TAMASHKDEDEKFPFSGVMIPEKE--SAGASNTALDNVMTDEBEIKATNPVATERFGT 579
 516 PVATTEDEKFPFSGVMIPEKE--SAGASNTALDNVMTDEBEIKATNPVATERFGT 575
 580 VAVNPFSSSTDPATGTHAMGALPGMWODRDVYLQGIPTWAKIPTHDGTHPSPLMGGRG 639
 576 TVTNBONTTAPPAELEVGLALPGMWODRDVYLQGIPTWAKIPTHDGTHPSPLMGGRG 635
 640 LKAPPOLILKTPVAPPAELEVGLALPGMWODRDVYLQGIPTWAKIPTHDGTHPSPLMGGRG 639
 636 LKAPPOLILKTPVAPPAELEVGLALPGMWODRDVYLQGIPTWAKIPTHDGTHPSPLMGGRG 635
 700 QYTSNATKASANDVPTVNNGLYTPRPIGRYLYRPL 736
 696 QYTSNATKASANDVPTVNNGLYTPRPIGRYLYRPL 732

RESULT 11

ID 067666 PRELIMINARY; PRT; 732 AA.
 AC 067666;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE VP1.
 GN VP1.
 OS goose parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 RN NCBI_Taxid=38251;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VIRULENT B;
 RX MEDLINE=96010229; PubMed=7571426;
 RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
 RT "Analysis of the complete nucleotide sequences of goose and muscovy
 duck parvoviruses indicates common ancestral origin with adeno-
 associated virus 2.";
 RL Virology 212:562-573 (1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=VIRULENT B;
 RA Zadori Z.; (Apr-1995) to the EMBL/GenBank/DBJ databases.
 RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U25749; AA83230.1; -
 DR Interpro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo coat; 1.
 SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 54.5%; Score 2173.5; DB 12; Length 732;
 Best Local Similarity 54.7%; Pred. No. 1.1e-136;
 Matches 414; Conservative 108; Mismatches 176; Indels 59; Gaps 14;

9 DMLNDLSGIRREWMWLPKGPAPKANKOQK-----DDGRGLVLPKYYLGPEN 57
 6 DSFEWYETLAASWRLKAGAPKPKSNQOSQSVSTRKPKORKNNGFVLPGYKYLGPEN 65
 58 GLDKGPRVNAADAALAHDKAYDQOLKAGDNPYLRYNHADAEPORLEOEDTSFGNLGRA 117
 66 GLDKGPRVNAADVALEHDKAYDQOLKAGDNPYIKFNHADOEDTIDNLSQDTSFGNLGRA 125
 118 VFOAKRRVLEPLGLVEGAKTAPGKRRPVQSGFOEPDSSGIGKTGQOPAKKRLNFQQTG 177
 126 VFOAKRRVLEPLGLVEGAKTAPGKRRPVQSGFOEPDSSGIGKTGQOPAKKRLNFQQTG 177
 169 KRLNFGQDSESVDPDPLG-----EPPTPAVGPPTMASGGGAPMADNNEGADGVGNASG 228
 185 -----GATN-----GTEP-----VAASEMAEGGAGGSSGAGGVGNASG 224
 229 HCDSTWLGADRYITTTSTRTWALPTYNHLYKQISSASTGASNDNHYFGYSTPWGFDEN 286
 225 HCDSTWLGADRYITTTSTRTWALPTYNHLYKQISSASTGASNDNHYFGYSTPWGFDEN 281
 287 RFHCSPSPDWORLNNNGFPRKRLNFKLNIQKVEVTNDGVTIANNLSTVOVPSD 346
 282 RFHCSPSPDWORLNNNGFPRKRLNFKLNIQKVEVTNDGVTIANNLSTVOVPSD 341
 347 SEYOLPYVIGSAGHOGCLPPFPADVEMIPQYGYLTN---NGSQAVERSSFYCLEYPPSQM 403
 342 DEHQLPYVIGSAGHOGCLPPFPADVEMIPQYGYLTN---NGSQAVERSSFYCLEYPPSQM 401
 404 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 463
 402 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 455
 402 LRTGNPFTEYTFEEVFPFSSVYAHSGSLDRLNMPILDQYLYNLRTONOGSGAQNKDLP 455
 464 SRGSPAGMSVQPKNMLPGPCYRQORVSKTK--TDNNNSFTWTGASKYNLNGESLIINGP 521
 456 KKAIVKAGAFGAMGNMLPGPCYRQORVSKTK--TDNNNSFTWTGASKYNLNGESLIINGP 515

QY 522 TAMASHKODEKFFPMGSGVMI FGKE--SAGASNTALDNVMTDEEIKATNPATERFGT 579
DB 516 PVSTVTEGEASSLPADNITLGIADOPYRSGSTAGISDINWTEOEVAFTNGVMKPYGR 575
QY 580 VAVFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGFG 639
DB 576 TVTNEQNTTTPATSSDDIVGALPGMWQNRDIYLOGPIMAKIPKTDGKHPHSPNLSGFG 635
QY 640 LKNPPOILIKNTVPANPPAEFSATKFASTIYQSTGVSVLEMELOKENSKRNNPEV 699
DB 636 LHNPPQVFKNTVPADPPVEYVHOKMNSYITQYSTGQCTVEWVWELRKENSKRNNPEI 695
QY 700 QYTSNVAKSANVDFTVDNNGLYTEPRPIGTRYLRPL 736
DB 696 QYTSNFSNRTSIMPAPNETGTYVEDRLIGTRYLTQNL 732

RESULT 12
ID 083289 PRELIMINARY; PRT: 732 AA.
AC 083289;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VP1 protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=37325;
RN [1]
RP SEQUENCE FROM N.A.
RA Zadori Z., Erdei J., Nagy J., Kiseri J.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X75093; CAAS2984.1; -;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81284 MW; 9FF6955BC15DF3AA CRC64;

Query Match 54.4%; Score 2168.5; DB 12; Length 732;
Best Local Similarity 55.2%; Pred. No. 2.3e-16;
Matches 418; Conservative 90; Mismatches 186; Indels 63; Gaps 13;

QY 9 DMLIEDNLSEGIKREWMDLKPGAPKPKANQOKO-----DDGRGLVLPGYKYLGPFN 57
DB 10 DWIE---TAAASRHLKAGAPKPKSNQOSQSVSTDKKPKKNNRGVLPGYKYLGPFN 65
QY 58 GLDKGEFVNAAADAALEHDKAYDQOLKAGDNPYLRYNHADAFOERLQEDTSGFGLGRA 117
DB 66 GLDKGPVNRKADSVALEHDKAYDQOLKAGDNPYIKFKHADQEFIDNLQDPTSGFGLGRA 125
QY 118 VFQAKKRVLEPLGLVEEGAKTAPGKKRPVEOSPOEPSSSGIGTGGQAPAKKRLNPGQTG 177
DB 126 VFQAKKRVLEPLGLVEEPVNTAPKK-----SSG-KLTDHPYKPKKLE-- 170
QY 178 DSESVPPPOPLG-----EPATPAVGPPTMASGGAPADNNEGADGVNAGS 226
DB 171 --ENSEPSPSSGSGASAAATBGSFVAP-----NMAEGSGMGSGAGADGVNAGS 222
QY 222 NMHCDSFWLGDRTVITSTRTVALPTYNHLYKOISSASTGASNDNHFGYSTPMGYDFN 286
DB 223 NMHCDSQMLGDTVITKTRTVALPSYNNHMYQALITSGTNPDSN--TQAGYSTPMGYDFN 281
QY 287 RFHCHSPRDMORLNNHMGFRPKRLNFKLNIQVKEVTNDGVTIANLTSVQVPSD 346
DB 282 RFHCHSPRDMORLNNHMGFRPKRLNFKLNIQVKEVTNDGVTIANLTSVQVPSD 341
QY 347 SEYOLPVLSAAGSCLPPFADVFMIPQVGYLTIN--NCSQAVGSSSFYCLEYFSSOM 403
DB 342 NEHOLPVLSATGTMPPPSDVTALPQYGYCTMHINOSAKRNDNSAFICYEFSSOM 401
QY 404 LRTGNNTFSYTFEEVPHSSVYASQSLDRIMNPLIDQVLYLNRTONOGSSAQNKLDF 463
DB 402 LRTGNNEFSFEEVPHSSVYASQSLDRIMNPLIDQVLYLNFSEV--NGGRNAQ-----F 455

QY 464 SRGSPAGMSVQPKWLPGPCYRQORVSKTK--TDNNNSNTPTWAGSKNNLNGRESTINPG 521
DB 456 KKAIVKAGAFAGMGNWLPGLPLDQVRVAASGGDNYANMSIWSKGNKVFLEKDEYLLQPG 515
QY 522 TAMASHKODEKFFPMGSGVMI FGKE--SAGASNTALDNVMTDEEIKATNPATERFGT 579
DB 516 PVATTHEDDASSVPANITIGAKDPRSGSTAGISDINWTEOEELAPTPNGVMKPYGL 575
QY 580 VAVFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGFG 639
DB 576 TVTNEQNTTTPATSSDDIVGALPGMWQNRDIYLOGPIMAKIPKTDGKHPHSPNLSGFG 635
QY 640 LKNPPOILIKNTVPANPPAEFSATKFASTIYQSTGVSVLEMELOKENSKRNNPEV 699
DB 636 LHNPPQVFKNTVPADPPVEYVHOKMNSYITQYSTGQCTVEWVWELRKENSKRNNPEI 695
QY 700 QYTSNVAKSANVDFTVDNNGLYTEPRPIGTRYLRPL 736
DB 696 QYTSNFSNRTSIMPAPNETGTYVEDRLIGTRYLTQNL 732

RESULT 13
ID 08V395 PRELIMINARY; PRT: 732 AA.
AC 08V395;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Capsid protein VP.
OS Goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=GPV-YG;
RA Ge Y., You Y., Xu Q.;
RT "Analysis of the major open reading frames' nucleotide sequences in
RT Goose parvovirus GPV-YG strain isolated in China."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF416726; AAL37722.1; -;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81456 MW; 73F2E4BC7659744B6 CRC64;

Query Match 54.3%; Score 2165.5; DB 12; Length 732;
Best Local Similarity 54.8%; Pred. No. 3.7e-136;
Matches 415; Conservative 101; Mismatches 182; Indels 59; Gaps 14;

QY 9 DMLIEDNLSEGIKREWMDLKPGAPKPKANQOKO-----DDGRGLVLPGYKYLGPFN 57
DB 6 DSEEWYETAAASWRNLKAGAPKPKNQOQSVPAREPERRRNSNGFVLPGYKYLGPFN 65
QY 58 GLDKGEFVNAAADAALEHDKAYDQOLKAGDNPYLRYNHADAFOERLQEDTSGFGLGRA 117
DB 66 GLDKGPVNRKADSVALEHDKAYDQOLKAGDNPYIKFKHADQEFIDNLQDPTSGFGLGRA 125
QY 118 VFQAKKRVLEPLGLVEEGAKTAPGKKR-----PVEOSPO-EPSSSGIGTGGQAPAK 168
DB 126 VFQAKKRVLEPLGLVEEDSINTAVKNTGKLTLDHYVAVKPKKLTVEVSAGGTSVAQDG- 184
QY 169 KRLNPGQTGSESVPPPOPLGEPATPAVGPPTMASGGAPADNNEGADGVNAGS 228
DB 185 -----GATVE-----GTEP-----VAAEEMEGGGGALGDASGGADGVNAGS 224
QY 229 HCDSTWLGDRVITSTRTVALPTYNHLYKOISSASTGASNDN--HYFGYSTPMGYDFN 286
DB 225 HCDSSQMGNTVITKTRTVALPSYNNHMYQALITSGTNPDSN--TQAGYSTPMGYDFN 281
QY 287 RFHCHSPRDMORLNNHMGFRPKRLNFKLNIQVKEVTNDGVTIANLTSVQVPSD 346
DB 282 RFHCHSPRDMORLNNHMGFRPKRLNFKLNIQVKEVTNDGVTIANLTSVQVPSD 341

QY 347 SEXOLPYVLGSAHQGLPPPADVPMIPOYGYLTN---NGSOAVGRSSFFCYCLEYFPGM 403
 DB 342 DEHQLPYVLGSAHQGLPPPADVPMIPOYGYLTNNGAARFNDRSATYCLEYFPGM 401
 QY 404 LRTGNPFSTYEEVFFSHSSVAHSQSLDRLMNPILDYLYLNTONOGSAGNKKDLF 463
 DB 402 LRTGNPFSTYEEVFFSHSSVAHSQSLDRLMNPILDYLYLNTONOGSAGNKKDLF 463
 QY 464 SRSPAGMSVQPKNMLPGPCYRQORV-SKTKTNNNSNFT-NTGASKYVNLNGESLIING 521
 DB 464 SRSPAGMSVQPKNMLPGPCYRQORV-SKTKTNNNSNFT-NTGASKYVNLNGESLIING 521
 QY 522 TAMASHKDEDEKFFPMGSGVIMFGKE--SAGASNTALDNVMTDEBEIKATNPATERFET 579
 DB 516 PVATHTTKVPAASIPANONITIGLAKDPYRSGSTTAGISDIDMTDEQVAPFNGVGMKPYG 575
 QY 580 VAVNFOSSTDPATGVDVHAMGALPGVWQDRVYLOGPIMAKI PHTDGHFHSPLMGFG 639
 DB 576 TVTNEQNTTAPFTSSDLIDVYALPGVWQDRVYLOGPIMAKI PHTDGHFHSPLMGFG 635
 QY 640 LKNPPQOLIKNTPPANPAEFSATKFAFTIOYSGOVSEIEMELCKENSKRNPEV 699
 DB 636 LKNPPQOLIKNTPPANPAEFSATKFAFTIOYSGOVSEIEMELCKENSKRNPEV 695
 QY 700 QYTSNYSKANSVDFVNDNGLYTEPRPIGRTYLRPL 736
 DB 696 QYTSNYSKANSVDFVNDNGLYTEPRPIGRTYLRPL 732
 RESULT 14
 ID 065444 PRELIMINARY; PRT; 732 AA.
 AC 065444
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Capsid protein VP.
 GN VP.
 OS Barbardie duck parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=39118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FM;
 RX MEDLINE=96010229; PubMed=7571426;
 RA Zadori Z., Stefancsik R., Rauch T., Klsary J.;
 RT "Analysis of the complete nucleotide sequences of goose and muscovy
 duck parvoviruses indicates common ancestral origin with adeno-
 RT associated virus 2."
 RL Virology 212:562-573(1995).
 DR EMBL; U22967; AAA83225.1;
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat.1.
 SQ SEQUENCE 732 AA; 81314 MW; D92360596E2D2C05 CRC64;
 Query Match 54.2%; Score 2162.5; DB 12; Length 732;
 Best Local Similarity 55.1%; Pred. No. 5.9e-136;
 Matches 417; Conservative 90; Mismatches 187; Indels 63; Gaps 13;
 QY 9 DWLENDLSEGIEMWDLKPGAPKPKANQOKO-----DDGRLVLPYGYKLGPPN 57
 DB 10 DWLENDLSEGIEMWDLKPGAPKPKANQOKO-----DDGRLVLPYGYKLGPPN 57
 QY 58 GLDKGPVNAADAALAEHDKAVDQOLKAGDNPYLRYNHDAEFQERLQEDTSFGNIGRA 117
 DB 66 GLDKGPVNAADAALAEHDKAVDQOLKAGDNPYLRYNHDAEFQERLQEDTSFGNIGRA 125
 QY 118 VPAKRRVLEPLGLVEGATTAAGKRRPVQSGQFPPSSSGIGTQGPAPKRLNGCGIG 177
 DB 126 VPAKRRVLEPLGLVEGATTAAGKRRPVQSGQFPPSSSGIGTQGPAPKRLNGCGIG 170
 QY 178 DSESVDPPQPLG-----EPATPAVGPPTTMASSGAPMADNNEGADGVGNASG 226

DB 171 --ENSPSPNSGGEASAAATGSEPAAP-----NMAEGSGAGMGDSAGADGVGNASG 222
 QY 227 NMHCDSWLGDRVITSTTALPTYNHLYKQISASISGASNDNHYFGYSTPMGFEDN 286
 DB 223 NMHCDSWLGDRVITSTTALPTYNHLYKQISASISGASNDNHYFGYSTPMGFEDN 281
 QY 287 RFHCSPSPDQWRLINNNNGRPPKRLNPLFNIOYKEVETNDQVTTIANLSTVQVPSD 346
 DB 282 RFHCSPSPDQWRLINNNNGRPPKRLNPLFNIOYKEVETNDQVTTIANLSTVQVPSD 341
 QY 347 SEXOLPYVLGSAHQGLPPPADVPMIPOYGYLTN---NGSOAVGRSSFFCYCLEYFPGM 403
 DB 342 DEHQLPYVLGSAHQGLPPPADVPMIPOYGYLTNNGAARFNDRSATYCLEYFPGM 401
 QY 404 LRTGNPFSTYEEVFFSHSSVAHSQSLDRLMNPILDYLYLNTONOGSAGNKKDLF 463
 DB 402 LRTGNPFSTYEEVFFSHSSVAHSQSLDRLMNPILDYLYLNTONOGSAGNKKDLF 463
 QY 464 SRSPAGMSVQPKNMLPGPCYRQORV-SKTKTNNNSNFT-NTGASKYVNLNGESLIING 521
 DB 464 SRSPAGMSVQPKNMLPGPCYRQORV-SKTKTNNNSNFT-NTGASKYVNLNGESLIING 521
 QY 522 TAMASHKDEDEKFFPMGSGVIMFGKE--SAGASNTALDNVMTDEBEIKATNPATERFET 579
 DB 516 PVATHTTKVPAASIPANONITIGLAKDPYRSGSTTAGISDIDMTDEQVAPFNGVGMKPYG 575
 QY 580 VAVNFOSSTDPATGVDVHAMGALPGVWQDRVYLOGPIMAKI PHTDGHFHSPLMGFG 639
 DB 576 TVTNEQNTTAPFTSSDLIDVYALPGVWQDRVYLOGPIMAKI PHTDGHFHSPLMGFG 635
 QY 640 LKNPPQOLIKNTPPANPAEFSATKFAFTIOYSGOVSEIEMELCKENSKRNPEV 699
 DB 636 LKNPPQOLIKNTPPANPAEFSATKFAFTIOYSGOVSEIEMELCKENSKRNPEV 695
 QY 700 QYTSNYSKANSVDFVNDNGLYTEPRPIGRTYLRPL 736
 DB 696 QYTSNYSKANSVDFVNDNGLYTEPRPIGRTYLRPL 732
 RESULT 15
 ID 067672 PRELIMINARY; PRT; 676 AA.
 AC 067672
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE VPI (Fragment).
 OS Goose parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=38251;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SHM 319;
 RX MEDLINE=95343541; PubMed=7618268;
 RA Brown K.E., Green S.W., Young N.S.;
 RT "Goose parvovirus--an autonomous member of the dependovirus genus?";
 RL Virology 210:283-291(1995).
 DR EMBL; U34761; AAA75286.1;
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat.1.
 FT NON TER 676
 SQ SEQUENCE 676 AA; 74800 MW; F28A086A115A417F CRC64;
 Query Match 49.9%; Score 1990.5; DB 12; Length 676;
 Best Local Similarity 54.6%; Pred. No. 1.6e-124;
 Matches 382; Conservative 95; Mismatches 163; Indels 59; Gaps 14;
 QY 9 DWLENDLSEGIEMWDLKPGAPKPKANQOKO-----DDGRLVLPYGYKLGPPN 57
 DB 6 DSEFEYETTAASWRLKAGAPQPKPNQOSQSVSDREBERKDNRRGFVLPYGYKLGPPN 65
 QY 58 GLDKGPVNAADAALAEHDKAVDQOLKAGDNPYLRYNHDAEFQERLQEDTSFGNIGRA 117

Db 66 GLDGPVNNKADSVLHDKAYDLQKAGDMPYIKFNHADDPIDSLQDDHSPFGNIGKA 125
QY 118 VFOAKKXVLEPLGLVEEGAKTAPGKGR-----PVEQSPQ-EPDSSSGTGKTOOPAK 168
Db 126 VFOAKKXVLEPLGLVEEGAKTAPGKGR-----PVEQSPQ-EPDSSSGTGKTOOPAK 184
QY 169 KRLNFGOTGDSSESVPDQPLGEPPATPAVGPPTMASGGGAPMADNNEGADGVGNASGNW 228
Db 185 -----GATAE-----GTEP-----VASEMAEGGAGMDSSGADGVGNASGNW 224
QY 229 HCDSTWLGDRVITSTRTWALPTYNNHLYKOISSASTGASNDNH--YFGYSTPMWGFDFN 286
Db 225 HCDSTWLGDRVITSTRTWALPTYNNHLYKOISSASTGASNDNH--YFGYSTPMWGFDFN 281
QY 287 RFHCHFSRDMQRLINNNMGFRPKLNFNIQVKEYTNDJVTIANNLTSTVOYFSD 346
Db 282 RFHCHFSRDMQRLINNNMGFRPKLNFNIQVKEYTNDJVTIANNLTSTVOYFSD 341
QY 347 SEYOLPYVLSAHGCLPFPADVEMTPOVGYLTIN--NGSOAVGRSSPYCLEYFSPOM 403
Db 342 DEHOLPYVLSAHGCLPFPADVEMTPOVGYLTIN--NGSOAVGRSSPYCLEYFSPOM 401
QY 404 LRTGNFTFSYTFEEVPHSSYASOSLDRILNPLIDQYLYLNRTONOGSSAQNKDLF 463
Db 402 LRTGNFTFSYTFEEVPHSSYASOSLDRILNPLIDQYLYLNRTONOGSSAQNKDLF 455
QY 464 SRGSPAGMSVOPKWLPGPCYRQORV-SKYKTNNNSNFT-WTGASKYMLNGRESIINPG 521
Db 456 KKAIVGAYGTWGRWMLPGPKFLDQRYVAVTGTIDNVANMNINSGNKVNLKDRQYLLQPG 515
QY 522 TAMASHDDEDEKFPMSGVMEFGKE--SAGASNTALDNVMTDEBEIKATNPVATERFGT 579
Db 516 PVSAAHTBGEASSIPAKNIIIGIADPYRSGSTTAGISIDIMWTDQEVAPNTGVGMKPYGR 575
QY 580 VAVNFOSSTDPATGDVYAMGALPGMWQDRDYYLQGPIMAKIPHTDGHFHPSPLMGFG 639
Db 576 TVTNEQNTTTPSTSDLDVGLALPGMWQDRDYYLQGPIMAKIPHTDGHFHPSPLMGFG 635
QY 640 LKAPPPQILLIKNTVPANPPAEFSATKPAFTIOYSTGQ 678
Db 636 LKAPPPQILLIKNTVPANPPAEFSATKPAFTIOYSTGQ 674

Search completed: July 17, 2003, 18:34:29
Job time : 58.7373 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 ; Search time 41.664 Seconds
(without alignments)
1915.734 Million cell updates/sec

Title: US-09-807-802a-15

Perfect score: 3251

Sequence: 1 TAPGKRPVQSPQEPDSSS.....NGLYTEPRPIGRVLTPL 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251	100.0	599	21	AAV71168 Adeno-associated v
2	3251	100.0	736	21	AAV71167 Adeno-associated v
3	3229	99.3	736	22	AAV59847 AAV6 capsid protel
4	2906	89.4	534	21	AAV71169 Adeno-associated v
5	2832	87.1	736	22	AAV59846 AAV3B capsid prote
6	2815	86.6	736	22	AAV59845 AAV3A capsid prote
7	2764.5	85.0	735	22	AAV5792 Adeno-associated v
8	2764.5	85.0	735	22	AAV51508 Adeno-associated v
9	2764.5	85.0	735	22	AAV59844 AAV2 capsid protel
10	2759.5	84.9	598	22	AAV65793 Adeno-associated v

11	2759.5	84.9	598	22	AAV51509	Adeno-associated v
12	2738	84.2	724	22	AAV50326	Adeno-associated v
13	2481.5	76.3	533	22	AAV65794	Adeno-associated v
14	2477.5	76.2	533	22	AAV51510	Adeno-associated v
15	1830.5	56.3	598	19	AAV46313	AAV4 VP2 coat prot
16	1830.5	56.3	734	19	AAV46308	AAV4 VP1 capsid pr
17	1719	52.9	736	18	AAV19000	Duck parvovirus ca
18	1701	52.3	732	16	AAV5385	Barbary duck parvo
19	1700.5	52.3	588	21	AAV58161	Adeno-associated v
20	1700.5	52.3	588	23	AAV11406	Adeno-associated v
21	1700.5	52.3	724	21	AAV58160	Adeno-associated v
22	1700.5	52.3	724	23	AAV11405	Adeno-associated v
23	1690.5	52.0	544	19	AAV46314	AAV4 VP3 coat prot
24	1676	51.6	534	16	AAV58386	Barbary duck parvo
25	1665	51.2	532	21	AAV58162	Adeno-associated v
26	1665	51.2	532	21	AAV11407	Adeno-associated v
27	772	23.7	756	21	AAV71231	Capsid protein enc
28	497.5	15.3	781	16	AAV08986	Human parvovirus V
29	497.5	15.3	781	20	AAV23227	Erythrovirus V9 VP
30	490	15.1	554	16	AAV08987	Human parvovirus V
31	476.5	15.1	554	20	AAV23230	Erythrovirus V9 VP
32	476.5	14.7	543	12	AAV13405	Parvo virus B19 VP
33	310	9.5	264	12	AAV13407	Parvo virus B19 PA
34	295.5	9.1	370	12	AAV13406	Parvo virus B19 PA
35	283	8.7	202	23	AAV83481	Adeno-associated v
36	250.5	7.7	686	5	AAV40068	Sequence of a porc
37	243.5	7.5	579	13	AAV29079	Porcine Parvovirus
38	216.5	6.7	579	14	AAV38697	PS1875 swine parvo
39	212.5	6.5	579	14	AAV38702	Swine parvovirus B
40	212.5	6.5	579	17	AAV99721	Swine parvovirus B
41	212.5	6.5	598	8	AAV70500	Pig parvo virus B
42	207.5	6.4	579	10	AAV94798	Swine parvovirus B
43	207	6.4	584	14	AAV30811	Feline parvovirus
44	203	6.2	472	5	AAV40675	Sequence encoded b
45	203	6.2	584	22	AAV04302	Feline panleukopen

ALIGNMENTS

RESULT 1
AAV71168
ID AAV71168 standard; Protein; 599 AA.
XX
AC AAV71168;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 capsid protein VP2.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; VP2.
XX
OS Adeno associated virus serotype 1.
XX
PN W0200028061.A2.
XX
PD 18-MAY-2000.
XX
PF 02-NOV-1999; 99WO-US25694.
XX
PR 05-NOV-1998; 98US-0107114.
XX
PA (VPE-) UNITV PENNSYLVANIA.
XX
PI Wilson JM, Xiao W;
XX
XX WPI: 2000-376571/32.
DR N-PSDB; AAD00772, AAD00778.
XX
PT Novel adeno-associated virus serotype 1 polynucleotide useful for
preparation of medicament for delivery of a transgene to a host

XX Claim 7; Page 93-95; 108pp; English.

PS The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA

CC which is characterized by two inverted terminal repeats (ITR) and open

CC reading frames for rep and capsid (cap) proteins. The rep reading frame

CC encodes four proteins, Rep 78, Rep 52 and Rep 40, while the cap

CC reading frame encodes three structural proteins, VP1, VP2 and VP3.

CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap

CC coding regions, are useful in production of recombinant viral vectors

CC for gene delivery. These vectors can be used as gene therapy

CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does

CC not induce the formation of neutralising antibodies specific to any

CC serotype of AAV hence is useful for transforming host cells, and in the

CC preparation of a medicament for the delivery of transgene to a host.

CC The present sequence is an AAV-1 cap protein VP2 which is

CC useful in the production of recombinant viral vector for gene delivery.

XX

SQ Sequence 599 AA;

Query Match 100.0%; Score 3251; DB 21; Length 599;

Best Local Similarity 100.0%; Pred. No. 1.3e-254;

Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 60

DB 1 TAPGKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 60

QY 61 VGPFTTASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTSRTTALPTYNHLY 120

DB 61 VGPFTTASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTSRTTALPTYNHLY 120

QY 121 KOISASTAGSANDNHHFGYSTPMGYDPDNRPHCHSPBDWQRLINNMWGRPRKLNFKLF 180

DB 121 KOISASTAGSANDNHHFGYSTPMGYDPDNRPHCHSPBDWQRLINNMWGRPRKLNFKLF 180

QY 181 NIQVKEVTTNDGVTITANNLTSTVQVFSDEYOLPYVLGSAHQCLPPPADVFMIPQYG 240

DB 181 NIQVKEVTTNDGVTITANNLTSTVQVFSDEYOLPYVLGSAHQCLPPPADVFMIPQYG 240

QY 241 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 300

DB 241 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 300

QY 241 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 300

DB 241 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 300

QY 301 LIIDQVLYLNRTQNOGSAONKDLFSRGSPPAGMSVOPKMWLPGPCYRQORVSKTKTDNN 360

DB 301 LIIDQVLYLNRTQNOGSAONKDLFSRGSPPAGMSVOPKMWLPGPCYRQORVSKTKTDNN 360

QY 361 NSNFTWTGASKYLNNGRESIINPGTAMASHKODEDKFFPMSCVMI FGKESAGASNTALDN 420

DB 361 NSNFTWTGASKYLNNGRESIINPGTAMASHKODEDKFFPMSCVMI FGKESAGASNTALDN 420

QY 421 VMTDEBERIKATNPVATRFCTVAVNFOSSSTDPATGVAHMGALPGKWODRVIYQGP 480

DB 421 VMTDEBERIKATNPVATRFCTVAVNFOSSSTDPATGVAHMGALPGKWODRVIYQGP 480

QY 481 IMAKIPIHTDGHFHSPLMGFGGLKNPPOQLIKATPVANPAEFSATKFSFTIOSTG 540

DB 481 IMAKIPIHTDGHFHSPLMGFGGLKNPPOQLIKATPVANPAEFSATKFSFTIOSTG 540

QY 541 QVSVEIEMELQKENSKRNNPEVOYTSYNAKSAANDFTVDNNGLYTEPRPIGRTYRLPPL 599

DB 541 QVSVEIEMELQKENSKRNNPEVOYTSYNAKSAANDFTVDNNGLYTEPRPIGRTYRLPPL 599

RESULT 2

AAV71167

ID AAV71167 standard; Protein; 736 AA.

XX

AC AAV71167;

XX

DT 08-SEP-2000 (first entry)

XX

DE Adeno-associated virus serotype 1 capsid protein VP1.

XX

KM Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;

KM cap protein; recombinant viral vector; gene delivery; gene therapy;

XX vaccine; transgene; VP1.

XX

OS Adeno associated virus serotype 1.

XX

PN MO200028061-A2.

XX

PD 18-MAY-2000.

XX

PF 02-NOV-1999; 99MO-US25694.

XX

PR 05-NOV-1998; 98US-0107114.

XX

PA (UYPB-) UNIV PENNSYLVANIA.

XX

PI Wilson JM, Xiao W;

XX

DR WPI: 2000-376571/32.

XX

DR N-PSDB: AAD00772, AAD00777.

XX

PT Novel adeno-associated virus serotype 1 polynucleotide useful for

PS preparation of medicament for delivery of a transgene to a host

XX

PS Claim 7; Page 87-90; 108pp; English.

XX

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA

CC which is characterized by two inverted terminal repeats (ITR) and open

CC reading frames for rep and capsid (cap) proteins. The rep reading frame

CC encodes four proteins, Rep 78, Rep 52 and Rep 40, while the cap

CC reading frame encodes three structural proteins, VP1, VP2 and VP3.

CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap

CC coding regions, are useful in production of recombinant viral vectors

CC for gene delivery. These vectors can be used as gene therapy

CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does

CC not induce the formation of neutralising antibodies specific to any

CC serotype of AAV hence is useful for transforming host cells, and in the

CC preparation of a medicament for the delivery of transgene to a host.

CC The present sequence is an AAV-1 cap protein VP1 which is

CC useful in the production of recombinant viral vector for gene delivery.

XX

SQ Sequence 736 AA;

Query Match 100.0%; Score 3251; DB 21; Length 736;

Best Local Similarity 100.0%; Pred. No. 1.7e-254;

Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 60

DB 138 TAPGKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 197

QY 61 VGPFTTASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTSRTTALPTYNHLY 120

DB 198 VGPFTTASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTSRTTALPTYNHLY 257

QY 121 KOISASTAGSANDNHHFGYSTPMGYDPDNRPHCHSPBDWQRLINNMWGRPRKLNFKLF 180

DB 258 KOISASTAGSANDNHHFGYSTPMGYDPDNRPHCHSPBDWQRLINNMWGRPRKLNFKLF 317

QY 181 NIQVKEVTTNDGVTITANNLTSTVQVFSDEYOLPYVLGSAHQCLPPPADVFMIPQYG 240

DB 318 NIQVKEVTTNDGVTITANNLTSTVQVFSDEYOLPYVLGSAHQCLPPPADVFMIPQYG 377

QY 241 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 300

DB 378 YLTINNGSQAVGRSSFYCLEFFPSQMLRTGNFTFSTFEVPPHSSYAHQSGLDRLNMP 437

QY 301 LIIDQVLYLNRTQNOGSAONKDLFSRGSPPAGMSVOPKMWLPGPCYRQORVSKTKTDNN 360

DB 438 LIIDQVLYLNRTQNOGSAONKDLFSRGSPPAGMSVOPKMWLPGPCYRQORVSKTKTDNN 497

```

QY 361 NSNFTWTGASKYMLNGRESIIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420
DB 498 NSNFTWTGASKYMLNGRESIIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557
QY 421 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRVYLQGP 480
DB 558 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRVYLQGP 617
QY 481 IMAKIPHTDGHFHPSPLMGGFGKLPKPPQILIKNTVPANPAPAEFSATKFASTTOYSTG 540
DB 618 IMAKIPHTDGHFHPSPLMGGFGKLPKPPQILIKNTVPANPAPAEFSATKFASTTOYSTG 677
QY 541 QVSVEIEMELQKENSARMNEVQYTSNYAKSANDFTVDNNGLYTERPRIGTRILTRPL 599
DB 678 QVSVEIEMELQKENSARMNEVQYTSNYAKSANDFTVDNNGLYTERPRIGTRILTRPL 736

RESULT 3
AAB59847
ID AAB59847 standard; Protein; 736 AA.
XX
AC AAB59847;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV6 capsid protein VP1.
XX
KM AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anemia; thalassemia;
XX blood clotting disorder; diabetes; capsid protein VP1.
OS Adeno associated virus.
XX
XX US6156303-A.
XX
XX 05-DEC-2000.
XX
XX 11-JUN-1997; 97US-0873168.
XX
XX 11-JUN-1997; 97US-0873168.
XX
XX (UNIV ) UNIV WASHINGTON.
XX
XX Russell DW, Rutledge EA;
XX
XX WPI; 2001-060164/07.
XX
XX Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX
XX Disclosure; Fig 2; 50pp; English.
XX
XX The present invention relates to adeno-associated virus serotypes. The
XX present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX
XX Sequence 736 AA;
XX
XX Query Match 99.3%; Score 3229; DB 22; Length 736;
XX Best Local Similarity 99.2%; Pred. No. 1e-252;
XX Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAPGKKRPVEQSPQEPDSSSGIGTKGQPAKKRINFQGTDSSEVPDPQIGEPATPAA 60
DB 138 TAPGKKRPVEQSPQEPDSSSGIGTKGQPAKKRINFQGTDSSEVPDPQIGEPATPAA 197

```

```

QY 61 VGPTTMAAGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTMALPTYNHLY 120
DB 198 VGPTTMAAGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTMALPTYNHLY 257
QY 121 KOISSASTGASNDNHFGYSTPWGYDFNRFCHSPSPDMQRIINNMGFRPLNPKLF 180
DB 258 KOISSASTGASNDNHFGYSTPWGYDFNRFCHSPSPDMQRIINNMGFRPLNPKLF 317
QY 181 NIOVKVTINDGVTIANNLTSVQVPSDEYOLPVLSAHQGCCLPPPADVFMIPOYG 240
DB 318 NIOVKVTINDGVTIANNLTSVQVPSDEYOLPVLSAHQGCCLPPPADVFMIPOYG 377
QY 241 YLTINNGSAVGRSSCYCLEYFPSSQMLRTGNFTFSYTEEYFHSYASOSLDIRLMP 300
DB 378 YLTINNGSAVGRSSCYCLEYFPSSQMLRTGNFTFSYTEEYFHSYASOSLDIRLMP 437
QY 301 LIDQYLYINRTQNGSAGAKNDLFSRGS PAGMSVQPKNMLPGCYRQORVSKYTTDNN 360
DB 438 LIDQYLYINRTQNGSAGAKNDLFSRGS PAGMSVQPKNMLPGCYRQORVSKYTTDNN 497
QY 361 NSNFTWTGASKYMLNGRESIIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420
DB 498 NSNFTWTGASKYMLNGRESIIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557
QY 421 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRVYLQGP 480
DB 558 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHMGALPGMWQDRVYLQGP 617
QY 481 IMAKIPHTDGHFHPSPLMGGFGKLPKPPQILIKNTVPANPAPAEFSATKFASTTOYSTG 540
DB 618 IMAKIPHTDGHFHPSPLMGGFGKLPKPPQILIKNTVPANPAPAEFSATKFASTTOYSTG 677
QY 541 QVSVEIEMELQKENSARMNEVQYTSNYAKSANDFTVDNNGLYTERPRIGTRILTRPL 599
DB 678 QVSVEIEMELQKENSARMNEVQYTSNYAKSANDFTVDNNGLYTERPRIGTRILTRPL 736

RESULT 4
AAY71169
ID AAY71169 standard; Protein; 534 AA.
XX
XX AAY71169;
XX
XX 08-SEP-2000 (first entry)
XX
XX Adeno-associated virus serotype 1 capsid protein VP3.
XX
XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX cap protein; recombinant viral vector; gene delivery; gene therapy;
XX vaccine; transgene; VP3.
XX
XX Adeno associated virus serotype 1.
XX
XX WO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX 02-NOV-1999; 99WO-US25694.
XX
XX 05-NOV-1998; 98US-0107114.
XX
XX (UNIV-) UNIV PENNSYLVANIA.
XX
XX Wilson DM, Xiao W;
XX
XX WPI; 2000-376571/32.
XX
XX N-PSDB; AAD00772, AAD00779.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
XX
XX Claim 7; Page 99-101; 108pp; English.
XX

```

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
CC which is characterised by two inverted terminal repeats (ITR) and open
CC reading frames for rep and capsid (cap) proteins. The rep reading frame
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
CC coding regions, are useful in production of recombinant viral vectors
CC for gene delivery. These vectors can be used as gene therapy
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC not induce the formation of neutralising antibodies specific to any
CC serotype of AAV hence is useful for transforming host cells, and in the
CC preparation of a medicament for the delivery of transgene to a host.
CC The present sequence is an AAV-1 cap protein VP3 which is
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 534 AA;

Query Match 89.4%; Score 2906; DB 21; Length 534;

Best Local Similarity 100.0%; Pred. No. 9,7e-227;

Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 66 MASGGGAPMADNNEGADGVNAGSGNWHCDSTWLGDRVITTSRTWALPTYNHLYKQISS 125
DB 1 MASGGGAPMADNNEGADGVNAGSGNWHCDSTWLGDRVITTSRTWALPTYNHLYKQISS 60
QY 126 ASTGASNDNHFGYSTPMGYDFENRFGHGFSPRDMQRLINNMGFRPKLNFKLNFQVK 185
DB 61 ASTGASNDNHFGYSTPMGYDFENRFGHGFSPRDMQRLINNMGFRPKLNFKLNFQVK 120
QY 186 EVTTNDGVTTIANNLTSTVOVFSDEYOLPYVLSAHOGLPPPADVFMIPQGYLTIN 245
DB 121 EVTTNDGVTTIANNLTSTVOVFSDEYOLPYVLSAHOGLPPPADVFMIPQGYLTIN 180
QY 246 NGSQAVGRSSVYCLEYFSPQMLRTGNMTFSYTFEEVPHSSVYAHOSLDLMLPLDQY 305
DB 181 NGSQAVGRSSVYCLEYFSPQMLRTGNMTFSYTFEEVPHSSVYAHOSLDLMLPLDQY 240
QY 306 LYIYNRTNOSGSAONKDLFSGRSPAGMSVOPKMWLGPCYRQORVSKTIDNNNSNFT 365
DB 241 LYIYNRTNOSGSAONKDLFSGRSPAGMSVOPKMWLGPCYRQORVSKTIDNNNSNFT 300
QY 366 WTGASKYNLNGRESIINGTAMASHKDEDEKFFPMGVMIFGKESAGASANTALDNMTD 425
DB 301 WTGASKYNLNGRESIINGTAMASHKDEDEKFFPMGVMIFGKESAGASANTALDNMTD 360
QY 426 EEEIKATNPVATERRFGTVAVNFOSSSTDPATGDVHAMGALPAMWQDDVYLQGPIMAKI 485
DB 361 EEEIKATNPVATERRFGTVAVNFOSSSTDPATGDVHAMGALPAMWQDDVYLQGPIMAKI 420
QY 486 PHTGHHFHPSPIMGFGGLKNPPOLLIKNTVPANPAPFASATKFASTITQYSTGOVSVE 545
DB 421 PHTGHHFHPSPIMGFGGLKNPPOLLIKNTVPANPAPFASATKFASTITQYSTGOVSVE 480
QY 546 IEMELQKENSKRMPDEVQYTSNAYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 599
DB 481 IEMELQKENSKRMPDEVQYTSNAYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 534
```

RESULT 5

AAB59846 ID AAB59846 standard; Protein; 736 AA.

XX AC AAB59846;

XX DT 28-MAR-2001 (first entry)

XX DE AAV3B capsid protein VP1.

XX AA3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

XX atherosclerosis; sickle cell anaemia; thalassemia;

XX KM blood clotting disorder; diabetes; capsid protein VP1.

XX OS Adeno associated virus.

```
XX XX US6156303-A.
XX XX 05-DEC-2000.
XX XX 11-JUN-1997; 97US-0873168.
XX XX 11-JUN-1997; 97US-0873168.
XX XX 11-JUN-1997; 97US-0873168.
XX XX (UNIV ) UNIV WASHINGTON.
XX XX Russell DW, Rutledge EA;
XX XX WPI; 2001-060164/07.
```

PT Adeno-associated virus serotype 6 and viral vector derived from it for

PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency

PT syndrome, sickle cell anemia, thalassemia and diabetes

PS Disclosure; Fig 2; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The
CC present sequence is capsid protein VP1 of one such serotype (AAV3B).
CC AAV3B can be used to construct AAV viral vectors for use in gene therapy
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
CC The AAV viral vectors have increased transduction efficiency of a
CC particular host cell as the AAV virion containing the AAV vector genome
CC can be modified to express a capsid protein of an AAV serotype that
CC transduces the selected host cell.

XX Sequence 736 AA;

Query Match 87.1%; Score 2832; DB 22; Length 736;

Best Local Similarity 85.7%; Pred. No. 1.5e-220;

Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;

```
QY 1 TARGKRPVQSPOEPDSSSGIGKTQQAQPAKRLNFGQTSDESVPDPOPLGPPATPAA 60
DB 138 TARGKRPVQSPOEPDSSSGIGKTQQAQPAKRLNFGQTSDESVPDPOPLGPPATPAA 137
QY 61 VGPITMASGGGAPMADNNEGADGVNAGSGNWHCDSTWLGDRVITTSRTWALPTYNHLY 120
DB 198 VGPITMASGGGAPMADNNEGADGVNAGSGNWHCDSTWLGDRVITTSRTWALPTYNHLY 257
QY 121 KQISSASTGASNDNHFGYSTPMGYDFENRFGHGFSPRDMQRLINNMGFRPKLNFKL 180
DB 258 KQISSASTGASNDNHFGYSTPMGYDFENRFGHGFSPRDMQRLINNMGFRPKLNFKL 120
QY 181 NIQVKEVTTNDGVTTIANNLTSTVOVFSDEYOLPYVLSAHOGLPPPADVFMIPQY 240
DB 317 NIQVKEVTTNDGVTTIANNLTSTVOVFSDEYOLPYVLSAHOGLPPPADVFMIPQY 316
QY 241 YLTINNGSAVGRSSVYCLEYFSPQMLRTGNMTFSYTFEEVPHSSVYAHOSLDLMLNP 300
DB 377 YLTINNGSAVGRSSVYCLEYFSPQMLRTGNMTFSYTFEEVPHSSVYAHOSLDLMLNP 436
QY 301 LIDQVLYLNRRTQ-NOSGSAONKDLFSGRSPAGMSVOPKMWLGPCYRQORVSKTIDN 359
DB 437 LIDQVLYLNRRTQ-NOSGSAONKDLFSGRSPAGMSVOPKMWLGPCYRQORVSKTIDN 436
QY 360 NNSNFTWTGASKYNLNGRESIINGTAMASHKDEDEKFFPMGVMIFGKESAGASANTALD 419
DB 497 NNSNFTWTGASKYNLNGRESIINGTAMASHKDEDEKFFPMGVMIFGKESAGASANTALD 419
QY 420 NVMTDEEBEIKATNPVATERRFGTVAVNFOSSSTDPATGDVHAMGALPAMWQDDVYLQ 479
DB 557 NVMTDEEBEIKATNPVATERRFGTVAVNFOSSSTDPATGDVHAMGALPAMWQDDVYLQ 479
QY 480 PIWAKIPHTDGHFHPSPIMGFGGLKNPPOLLIKNTVPANPAPFASATKFASTITQYST 539
DB 617 PIWAKIPHTDGHFHPSPIMGFGGLKNPPOLLIKNTVPANPAPFASATKFASTITQYST 539
```

QY 540 GQSVSEIEMELQKNSKRMNPEVQYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLRL 599
 DB 677 GQSVSEIEMELQKNSKRMNPEIQYTSNYSKASVAVDFTVDNNGLYSEPRPIGTRYLRL 736

RESULT 6
 ID AAB59845 standard; Protein: 736 AA.

XX AAB59845;

XX 28-MAR-2001 (first entry)

XX AAV3A capsid protein VP1.

XX AAV3A, gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

XX Atherosclerosis; sickle cell anaemia; thalassemia;

XX blood clotting disorder; diabetes; capsid protein VP1.

XX Adeno associated virus.

XX US6156303-A.

XX 05-DEC-2000.

XX 11-JUN-1997; 97US-0873168.

XX 11-JUN-1997; 97US-0873168.

XX (UNIW) UNIV WASHINGTON.

XX Russell DW, Rutledge EA;

XX WPI; 2001-060164/07.

XX Adeno-associated virus serotype 6 and viral vector derived from it for

XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency

XX syndrome, sickle cell anemia, thalassemia and diabetes

XX Disclosure; Fig 2; 50pp; English.

XX The present invention relates to adeno-associated virus serotypes. The

XX present sequence is capsid protein VP1 of one such serotype (AAV3A).

XX AAV3A can be used to construct AAV viral vectors for use in gene therapy

XX for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,

XX sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.

XX The AAV viral vectors have increased transduction efficiency of a

XX particular host cell as the AAV virion containing the AAV vector genome

XX can be modified to express a capsid protein of an AAV serotype that

XX transduces the selected host cell.

XX Sequence 736 AA;

XX Query Match 86.6%; Score 2815; DB 22; Length 736;

XX Best Local Similarity 85.2%; Pred. No. 3.6e-219;

XX Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;

QY 1 TAREKRPVEOSPOEPDSSGIGCTGQOPAKRLNFCQGTDSBSPPOPLGEPAPAA 60
 DB 138 TAREKKAIVDQSPDSSGIGCTGQOPAKRLNFCQGTDSBSPPOPLGEPAPAA 197
 QY 61 VGPPTMASGGAPADNNEGADGVNAGNMGHCHSTLGDRTVITSTRTALPTNNHLY 120
 DB 198 LGSNTMASGGAPADNNEGADGVNAGNMGHCHSTLGDRTVITSTRTALPTNNHLY 257
 QY 121 KOISSASTGASNDNHFGYSTPWCYFPFNRHCHFSRDMQRLINNMGFRPKLNF 180
 DB 258 KOISSQS-GASNDNHFGYSTPWCYFPFNRHCHFSRDMQRLINNMGFRPKLNF 316
 QY 181 NIQVKEVTTDGVTTIANLSTYQVPSDSRYQPLVLSAHQCLPPFPADVMIQYX 240
 DB 317 NIQVKEVTTDGVTTIANLSTYQVPSDSRYQPLVLSAHQCLPPFPADVMIQYX 376

QY 241 YLTINNGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTEEVPFHSSYAHSSQSLDLMLNP 300
 DB 377 YLTINNGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTEEVPFHSSYAHSSQSLDLMLNP 436
 QY 301 LIDQVLYLNRTQ-NQSGSAQNKDLLFSRGSFAGMSVQPKNMLPGCYROQVSKTKTDN 359
 DB 437 LIDQVLYLNRTQGTGTTNQGRLFFSQAGPQMSLQANMLPGCYROQVSKTKTDN 496
 QY 360 NNSNFTMGASKYHLNGRBSLTPGATMA SHKDEDEKFPFMSGVMI FGKESAGASATLAD 419
 DB 497 NNSNFTMGASKYHLNGRBSLTPGATMA SHKDEDEKFPFMSGVMI FGKESAGASATLAD 556
 QY 420 NMTTDEEIKATNPVATERFGVAVAVPOSSSTDPATGVDYHAGALPGMWQDRDYLQ 479
 DB 557 NMTTDEEIKATNPVATERFGVAVAVPOSSSTDPATGVDYHAGALPGMWQDRDYLQ 616
 QY 480 PIWAKIPHTDGHFHPSPMLMGFGLKHPPOILIKNTVPANPAEPAATFQYST 539
 DB 617 PIWAKIPHTDGHFHPSPMLMGFGLKHPPOILIKNTVPANPAEPAATFQYST 676
 QY 540 GQSVSEIEMELQKNSKRMNPEVQYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLRL 599
 DB 677 GQSVSEIEMELQKNSKRMNPEIQYTSNYSKASVAVDFTVDNNGLYSEPRPIGTRYLRL 736

RESULT 7

ID AAG65792 standard; Protein: 735 AA.

XX AAG65792;

XX 11-FEB-2002 (first entry)

XX Adeno-associated virus 2 (AAV-2) major coat protein VP1.

XX Recombinant viral vector; RVV; capsid; parvovirus; transgenic; cytosolic;

XX inverted terminal repeat; neuroprotective; antineoplastic; ITR;

XX antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;

XX major coat protein; AAV-2; VP1.

XX Adeno-associated virus 2.

XX WO200168888-A2.

XX 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07927.

XX 14-MAR-2000; 2000US-189110P.

XX (NEUR-) NEUROLOGIX INC.

XX Xiao W, During MJ;

XX WPI; 2001-596912/67.

XX N-PADB; AAI66974.

XX Recombinant viral vector useful in improving gene therapy in a subject,

XX and for increasing efficiency of entry into a cell, comprises a

XX transgene -

XX Disclosure; Page 50; 53pp; English.

XX The invention provides a recombinant viral vector (RVV) comprising a

XX chimeric capsid (I) having at least one non-native amino acid sequence,

XX derived from a capsid protein domain of parvovirus (II), a virus (III),

XX or their combination, and a transgene flanked 5' and 3' by inverted

XX terminal repeat (ITR) sequences, derived from (II), (III), or their

XX combination. The RVV is useful for improving gene therapy in a subject

XX with a disorder, and for increasing the efficiency of entry into a cell,

XX which involves producing (I) encapsulating a viral vector, and contacting

XX a cell with RVV having (I) such that (I) binds to an attachment site on

CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (i) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Leach-Nyhan syndrome, amyloid polynuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include,
CC tumours especially central nervous system tumours, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences
CC are used in the construction of a chimeric vector.

XX Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;

Best Local Similarity 82.8%; Pred. No. 4.4e-215;

Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKRPVPSQFQEPSSSGIGKTGQOPAKKRLNFQGTGSESVPPQPLGEPATPAA 60

DB 138 TAPGKRPVPSQFQEPSSSGIGKTGQOPAKKRLNFQGTGSESVPPQPLGEPATPAA 197

QY 61 VGPPTMASGGAPADNNEGADGVGNAGNHCSTWMLGDRVITTSRTTALPTYNNHLY 120

DB 198 LGTMTATGSAAPADNNEGADGVGNAGNHCSTWMLGDRVITTSRTTALPTYNNHLY 257

QY 121 KOISSASTGASNDNHFGYSTPWGYPDRNFCHSPRDWQRLINNMGFRPKLNFPLF 180

DB 258 KOISSAS-GASNDNHFGYSTPWGYPDRNFCHSPRDWQRLINNMGFRPKLNFPLF 316

QY 181 NIQVKEVTTNDGVTIANNTLSTVOVPSDSEYOLPYVLGSAHOGCLPPFPADVFMVPOYG 240

DB 317 NIQVKEVTTNDGVTIANNTLSTVOVPSDSEYOLPYVLGSAHOGCLPPFPADVFMVPOYG 376

QY 241 YLTNNGSQAVGRSSPYCLEYFPSPQMLRTGNNFTFSYFEEVPHSHYASOSLDRLMNP 300

DB 377 YLTNNGSQAVGRSSPYCLEYFPSPQMLRTGNNFTFSYFEEVPHSHYASOSLDRLMNP 436

QY 301 LIDQVLYLNTNONGSAONKDLFSPRGSAPGMSVOPKXMLPGPCYRQORVSKTKTDNN 360

DB 437 LIDQVLYLNTNONGSAONKDLFSPRGSAPGMSVOPKXMLPGPCYRQORVSKTKTDNN 496

QY 361 NSNFTWGTASKYXNNGRETIINPGTAMASHKDEDEKFPFGSGVLIIPGQSEKTNVIEK 556

DB 497 NSNFTWGTASKYXNNGRETIINPGTAMASHKDEDEKFPFGSGVLIIPGQSEKTNVIEK 556

QY 421 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVAHMGALPGMWQDDRVYLOGP 480

DB 557 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVAHMGALPGMWQDDRVYLOGP 616

QY 481 IWAKIPHTDGHFSPPLMGFGGLKXNPPOILLIKNTVPANPAPBSATKPSFTTOYSTG 540

DB 617 IWAKIPHTDGHFSPPLMGFGGLKXNPPOILLIKNTVPANPAPBSATKPSFTTOYSTG 676

QY 541 QVSVEIEMELQKENSXKRNPEVOYTSNYSKASANDFTVNNGLYTEPRPIGRIYLRPL 599

DB 677 QVSVEIEMELQKENSXKRNPEVOYTSNYSKASANDFTVNNGLYTEPRPIGRIYLRPL 735

RESULT 8

ID AAM51508 standard; protein; 735 AA.

XX AAM51508;

XX 02-JAN-2002 (first entry)

DE Adeno-associated virus VP1 capsid protein.

XX Adeno-associated virus; AAV; VP1; capsid; virus-like particle;

KW nuclear localisation signal; VP3.

XX Adeno associated virus.

XX JP2001169777-A.

XX 26-JUN-2001.

XX 30-JUL-1999; 990P-0249140.

XX 30-JUL-1999; 990P-0249140.

XX 30-JUL-1999; 990P-0249140.

XX (HANA) HANA H.

XX WPI; 2001-599854/68.

XX New virus-like particles from VP3 capsid protein of adeno-associated

XX virus, comprise a peptide containing a nucleus-shifting signal

XX connected to its N-terminal -

XX Disclosure; Page 10-13; 33pp; Japanese.

XX The present sequence is provided in a specification relating to a

XX virus-like particle-forming protein, and to a peptide containing a

XX nuclear-shifting signal at its N-terminus that can form virus-like

XX particles by shifting to the nucleus of the animal cell in which it

XX is expressed. The method is used for forming virus-like particles

XX from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;

Best Local Similarity 82.8%; Pred. No. 4.4e-215;

Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKRPVPSQFQEPSSSGIGKTGQOPAKKRLNFQGTGSESVPPQPLGEPATPAA 60

DB 138 TAPGKRPVPSQFQEPSSSGIGKTGQOPAKKRLNFQGTGSESVPPQPLGEPATPAA 197

QY 61 VGPPTMASGGAPADNNEGADGVGNAGNHCSTWMLGDRVITTSRTTALPTYNNHLY 120

DB 198 LGTMTATGSAAPADNNEGADGVGNAGNHCSTWMLGDRVITTSRTTALPTYNNHLY 257

QY 121 KOISSASTGASNDNHFGYSTPWGYPDRNFCHSPRDWQRLINNMGFRPKLNFPLF 180

DB 258 KOISSAS-GASNDNHFGYSTPWGYPDRNFCHSPRDWQRLINNMGFRPKLNFPLF 316

QY 181 NIQVKEVTTNDGVTIANNTLSTVOVPSDSEYOLPYVLGSAHOGCLPPFPADVFMVPOYG 240

DB 317 NIQVKEVTTNDGVTIANNTLSTVOVPSDSEYOLPYVLGSAHOGCLPPFPADVFMVPOYG 376

QY 241 YLTNNGSQAVGRSSPYCLEYFPSPQMLRTGNNFTFSYFEEVPHSHYASOSLDRLMNP 300

DB 377 YLTNNGSQAVGRSSPYCLEYFPSPQMLRTGNNFTFSYFEEVPHSHYASOSLDRLMNP 436

QY 301 LIDQVLYLNTNONGSAONKDLFSPRGSAPGMSVOPKXMLPGPCYRQORVSKTKTDNN 360

DB 437 LIDQVLYLNTNONGSAONKDLFSPRGSAPGMSVOPKXMLPGPCYRQORVSKTKTDNN 496

QY 361 NSNFTWGTASKYXNNGRETIINPGTAMASHKDEDEKFPFGSGVLIIPGQSEKTNVIEK 556

DB 497 NSNFTWGTASKYXNNGRETIINPGTAMASHKDEDEKFPFGSGVLIIPGQSEKTNVIEK 556

QY 421 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVAHMGALPGMWQDDRVYLOGP 480

DB 557 VMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVAHMGALPGMWQDDRVYLOGP 616

QY 481 IWAKIPHTDGHFSPPLMGFGGLKXNPPOILLIKNTVPANPAPBSATKPSFTTOYSTG 540

DB 617 IWAKIPHTDGHFSPPLMGFGGLKXNPPOILLIKNTVPANPAPBSATKPSFTTOYSTG 676

QY 541 QVSVEIEMELQKENSXKRNPEVOYTSNYSKASANDFTVNNGLYTEPRPIGRIYLRPL 599

DB 677 QVSVEIEMELQKENSXKRNPEVOYTSNYSKASANDFTVNNGLYTEPRPIGRIYLRPL 735

CC which involves producing (1) encapsulating a viral vector, and contacting
CC a cell with RVV having (1) such that (1) binds to an attachment site on
CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (1) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Leisch-Nyhan syndrome, amyloid polynuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include
CC tumours especially central nervous system tumours, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences
CC are used in the construction of a chimeric vector.

SQ Sequence 598 AA;

Query Match 84.9%; Score 2759.5; DB 22; Length 598;
Best Local Similarity 82.8%; Pred. No. 8.4e-215;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFQGTGSESVDPDPLGEPATPAAY 61
DB 2 APGKKRPVEHSPVEPDSSSGTGAKGQOPAKKRLNFGQTDADSVDPDPLGEPAPASGL 61
QY 62 GPTTMAAGGAPADNNEGAGVGNASGNWHDSTWLGDRVITTSRTWALPTNNHLTK 121
DB 62 GTTMTATGSGAPADNNEGAGVGNSSGNWHDSTWMDRIVITTSRTWALPTNNHLTK 121
QY 122 QISSASTGASNDNHYFGYSTPMGYFDPNRFCHFSRPMQRLINNNWGFPRKLNFLKN 181
DB 122 QISSQS-GASNDNHYFGYSTPMGYFDPNRFCHFSRPMQRLINNNWGFPRKLNFLKN 180
QY 182 IQVEKVTNDGVTIANNLSTVOVFSDESEYQLPYVLGSAHQCLPPPADVEMIPQYGY 241
DB 181 IQVEKVTNDGVTIANNLSTVOVFTDSEYQLPYVLGSAHQCLPPPADVEMIPQYGY 240
QY 242 LTTNNGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPVHSSVAHSQSLDRMLNPL 301
DB 241 LTTNNGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPVHSSVAHSQSLDRMLNPL 300
QY 302 IDQYLYLNTQNGSAGNOKDLFSGRSFAGMSVQPKMLFGPCYRQOQVSKTKTDNNN 361
DB 301 IDQYLYLNTQNGSAGNOKDLFSGRSFAGMSVQPKMLFGPCYRQOQVSKTKTDNNN 360
QY 362 SNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMI FGKESAGASNTALDNV 421
DB 361 SEYSMTGATKYHLNGRDLVNPGRAMASHKDEDEKFFQSGVLIFGKQSEKTNVIEKV 420
QY 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDNHAGALPGWVQDRDVLQGP 481
DB 421 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDNHAGALPGWVQDRDVLQGP 480
QY 482 WAKIPHTDGHFHPPLMGFGGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 541
DB 481 WAKIPHTDGHFHPPLMGFGGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 540
QY 542 VSVSEIWELOKENSKNMNEVOYTSNYSKASANDVFTVNNNGIYTEPRPIGRVYLRPL 599
DB 541 VSVSEIWELOKENSKNMNEVOYTSNYSKASANDVFTVNNNGIYTEPRPIGRVYLRPL 598

RESULT 11

AAM51509 standard; protein, 598 AA.

AC AAM51509;

XX 02-JAN-2002 (first entry)

XX Adeno-associated virus VP2 capsid protein.

XX Adeno-associated virus; AAV; VP2; capsid; virus-like particle;
KW nuclear localisation signal; VP3.
XX Adeno associated virus.
OS JP2001169777-A.PN 26-JUN-2001.PD 30-JUL-1999; 99UD-0249140.
PF 30-JUL-1999; 99UD-0249140.
PR 30-JUL-1999; 99UD-0249140.
XX (HAND/) HANDA H.
PA WPI; 2001-599854/68.
DR WPI; 2001-599854/68.
XX New virus-like particles from VP3 capsid protein of adeno-associated
PT virus, comprise a peptide containing a nucleus-shifting signal
PT connected to its N-terminal -
PS Claim 1, Page 14-16; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a
CC virus-like particle-forming protein, and to a peptide containing a
CC nuclear-shifting signal at its N-terminus that can form virus-like
CC particles by shifting to the nucleus of the animal cell in which it
CC is expressed. The method is used for forming virus-like particles
CC from the VP3 capsid protein of adeno-associated virus (AAV).

SQ Sequence 598 AA;

Query Match 84.9%; Score 2759.5; DB 22; Length 598;
Best Local Similarity 82.8%; Pred. No. 8.4e-215;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFQGTGSESVDPDPLGEPATPAAY 61
DB 2 APGKKRPVEHSPVEPDSSSGTGAKGQOPAKKRLNFGQTDADSVDPDPLGEPAPASGL 61
QY 62 GPTTMAAGGAPADNNEGAGVGNASGNWHDSTWLGDRVITTSRTWALPTNNHLTK 121
DB 62 GTTMTATGSGAPADNNEGAGVGNSSGNWHDSTWMDRIVITTSRTWALPTNNHLTK 121
QY 122 QISSASTGASNDNHYFGYSTPMGYFDPNRFCHFSRPMQRLINNNWGFPRKLNFLKN 181
DB 122 QISSQS-GASNDNHYFGYSTPMGYFDPNRFCHFSRPMQRLINNNWGFPRKLNFLKN 180
QY 182 IQVEKVTNDGVTIANNLSTVOVFSDESEYQLPYVLGSAHQCLPPPADVEMIPQYGY 241
DB 181 IQVEKVTNDGVTIANNLSTVOVFTDSEYQLPYVLGSAHQCLPPPADVEMIPQYGY 240
QY 242 LTTNNGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPVHSSVAHSQSLDRMLNPL 301
DB 241 LTTNNGSOAVGRSSFYCLEYFPQMLRTGNFTFSYTFEVPVHSSVAHSQSLDRMLNPL 300
QY 302 IDQYLYLNTQNGSAGNOKDLFSGRSFAGMSVQPKMLFGPCYRQOQVSKTKTDNNN 361
DB 301 IDQYLYLNTQNGSAGNOKDLFSGRSFAGMSVQPKMLFGPCYRQOQVSKTKTDNNN 360
QY 362 SNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMI FGKESAGASNTALDNV 421
DB 361 SEYSMTGATKYHLNGRDLVNPGRAMASHKDEDEKFFQSGVLIFGKQSEKTNVIEKV 420
QY 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDNHAGALPGWVQDRDVLQGP 481
DB 421 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDNHAGALPGWVQDRDVLQGP 480
QY 482 WAKIPHTDGHFHPPLMGFGGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 541
DB 481 WAKIPHTDGHFHPPLMGFGGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 540

QY 542 VSVEIEMELOKENSKRNNPEVOYTSNYAKSANVDFVDDNGLYTEPRIGTRYLTRPL 599
DB 541 VSVEIEMELOKENSKRNNPEIQYTSNYSKSVNDFVDDNGLYTEPRIGTRYLTRPL 598

RESULT 12
ID AAB50326 standard; protein; 734 AA.
XX AAB50326;
AC AAB50326;
XX
DT 09-MAR-2001 (first entry)
XX
DE Adeno-associated virus capsid protein sequence.
XX
KM Adeno-associated virus; AAV; capsid; virus binding inhibition;
XX competitive inhibitor.
XX Mastadenovirus.
OS Synthetic.
XX MO200073316-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000MO-US144466.
XX
PR 28-MAY-1999; 99US-0321589.
XX
PA (CELL-) CELL GENESYS INC.
XX
PI Patel S, McArthur J;
XX
DR WPI; 2001-061507/07.
XX
PT New polypeptide portion of a virus protein binding to an antibody
XX specific for the virus useful for inhibiting binding of the virus to a
PT cell or for binding host antibody to provide a transient tolerant or
PT non-responsive state -
XX
PS disclosure; Fig 2; 33pp; English.
XX
CC The present sequence is given in a specification relating to
CC polypeptide portions of a virus protein or its derivative, that bind to
CC an antibody specific for the virus or inhibit binding of the virus to a
CC cell. The polypeptides are used in inhibiting the binding of viruses to
CC cells of a host. Oligopeptides that inhibit binding of virus to
CC receptor can be used as competitive inhibitors to release bound virus
CC in an adsorption-type assay, and if an antibody was used as an
CC immunoadsorbent, the oligopeptide could be used to elute bound virus
CC from a solid support to which virus antibody is immobilised. These
CC oligopeptides may further be used to bind to host antibody to provide a
CC transient tolerant or non-responsive state.
XX
SQ Sequence 734 AA;

Query Match 84.2%; Score 2738; DB 22; Length 734;
Best Local Similarity 82.5%; Pred. No. 6.2e-213;
Matches 494; Conservative 43; Mismatches 60; Indels 2; Gaps 2;

QY 1 TABGKRPEVQSOEPDSSGIGTGQOPAKRRLNFQOTGSDSSVDPDPOPLGEPPTPA 60
DB 138 TABGKRPEVQSOEPDSSGIGTGQOPAKRRLNFQOTGSDSSVDPDPOPLGEPPTPA 60
DB 198 LGTNMATSGADPADNNEGADGVGNAGNHCDSITWMDRVITTSRTWALPTYNHLY 257
QY 61 VGEPTNAGGAGMADNNEGADGVGNAGNHCDSITWMDRVITTSRTWALPTYNHLY 120
DB 198 LGTNMATSGADPADNNEGADGVGNAGNHCDSITWMDRVITTSRTWALPTYNHLY 257
QY 121 KOISSASTAGSNHNYGYSTPMGYPFRPNFCHFSRDMQRLINNNGRPRKLNFKLP 180
DB 258 KOISSOS-GASNNHNYGYSTPMGYPFRPNFCHFSRDMQRLINNNGRPRKLNFKLP 316
QY 181 NIQKEVTTNDVTTIANNLSTVQVPSDSEYOLPYLGAHQGLPPFADVFMTPOYG 240

DB 317 NIQKEVTTNDVTTIANNLSTVQVPSDSEYOLPYLGAHQGLPPFADVFMTPOYG 376
QY 241 YLTINNGSQAVGSSFCLEYPFSSQMLRTGNNTFSTYEEVPHSSYAHOSLDRLAMP 300
DB 377 YLTINNGSQAVGSSFCLEYPFSSQMLRTGNNTFSTYEEVPHSSYAHOSLDRLAMP 436
QY 301 LIDQVLYLNRTOGNSAONKDLFSSRGS PGMGSVQPKMVLPGCYRQORVSKTDDNN 360
DB 437 LIDQVLYLNRTOGNSAONKDLFSSRGS PGMGSVQPKMVLPGCYRQORVSKTDDNN 496
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNLALDN 420
DB 497 NSEYSWTGAFTKHLNGRDGLVNPGRAMASHKDEDEKFPMSGVLIRGQSEKTNVDIEK 556
QY 421 VMTDEEIKATNPVATFERPGYAVNVFGSSSTDPAAGDVHANGALFGMWQDRDYLOGP 480
DB 557 VMTDEEIKATNPVATFERPGYAVNVFGSSSTDPAAGDVHANGALFGMWQDRDYLOGP 615
QY 481 IMAKIPHTDGHFSPFLMGSGFGLKNPPOILLKNTVPANPAPFSAATKPAFPIYOSTG 540
DB 616 IMAKIPHTDGHFSPFLMGSGFGLKNPPOILLKNTVPANPAPFSAATKPAFPIYOSTG 675

QY 541 QVSVEIEMELOKENSKRNNPEVOYTSNYAKSANVDFVDDNGLYTEPRIGTRYLTRPL 599
DB 676 QVSVEIEMELOKENSKRNNPEIQYTSNYSKSVNDFVDDNGLYTEPRIGTRYLTRPL 734

RESULT 13
ID AAG65794 standard; protein; 533 AA.
XX AAG65794;
AC AAG65794;
XX
DT 11-FEB-2002 (first entry)
XX
DE Adeno-associated virus 2 (AAV-2) major coat protein VP3.
XX
KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
KM inverted terminal repeat; nonretroviral; neuroprotective; anti-neuritic; ITR;
KM antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
XX major coat protein; AAV-2; VP3.
XX
OS Adeno-associated virus 2.
XX
PN MO200168888-A2.
XX
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001MO-US07927.
XX
PR 14-MAR-2000; 2000US-189110P.
XX
PA (NEUR-) NEUROLOGIX INC.
XX
PI Xiao W, During MJ;
XX
DR WPI; 2001-596912/67.
XX
PT N-PSDB; AA166974.
XX
PT Recombinant viral vector useful in improving gene therapy in a subject,
PT and for increasing efficiency of entry into a cell, comprises a desired
PT chimeric capsid having one non-native amino acid sequence and a desired
PT transgene -
XX
PS disclosure; Page 51; 53pp; English.
XX
CC The invention provides a recombinant viral vector (RVV) comprising a
CC chimeric capsid (I) having at least one non-native amino acid sequence,
CC derived from a capsid protein domain of parvovirus (II), a virus (III),
CC or their combination, and a transgene flanked 5' and 3' by inverted
CC terminal repeat (ITR) sequences, derived from (II), (III), or their
CC combination. The RVV is useful for improving gene therapy in a subject

CC with a disorder, and for increasing the efficiency of entry into a cell,
CC which involves producing (I) encapsulating a viral vector, and contacting
CC a cell with RVV having (I) such that (I) binds to an attachment site on
CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (I) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include
CC tumours especially central nervous system tumours, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences
CC are used in the construction of a chimeric vector.

SQ Sequence 533 AA;

Query Match 76.3%; Score 2481.5; DB 22; Length 533;

Best Local Similarity 83.3%; Pred. No. 2.4e-192; Mismatches 1; Gaps 1;

Matches 445; Conservative 37; Indels 1; Gaps 1;

66 MASGCGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTYNNHLYKQISS 125
1 MATSGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTYNNHLYKQISS 60
126 ASTGASDNHNYFGYSTPWGTFDNRFHCHSPRDMORLNNNGFPRKLNFLKFNIOVK 185
61 QS-GASNDNHYFGYSTPWGTFDNRFHCHSPRDMORLNNNGFPRKLNFLKFNIOVK 119
186 EVTTNDGVTITANNLSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMVPOGYLTIN 245
120 EVTQNDGVTITANNLSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMVPOGYLTIN 179
246 NGGQAVGRSSFFCLEYFPSSQMLRTGNNFTSYTFEEVFFHSSVAHSQSLDRLNPLIDQY 305
180 NGGQAVGRSSFFCLEYFPSSQMLRTGNNFTSYTFEEVFFHSSVAHSQSLDRLNPLIDQY 239
306 LYYINRTQNGSGAONKDLFSRGSPPAGMSVQPKMLPGPCYQORVSKTKTDNNNSNFT 365
240 LYYISRTNTPSGITTSRLOFSQAGASDIRDSRNLPGPCYQORVSKTKTDNNNSNFT 299
366 WTGASKTNLNGRESIINPGTAMASHKDEDEKFFPMGSGVMI FGKESAGSNALDNVITD 425
300 WTGATKTHLNGRSLVNPGRPMASHKDEDEKFFPMGSGVMI FGKESAGSNALDNVITD 359
426 EEEIKATNPVATERFGTVAVNFPSSSTDPATGVHAGMALPGVWMDRDVYLQGPWAKI 485
360 EEEIRTTNPVATERFGTVAVNFPSSSTDPATGVHAGMALPGVWMDRDVYLQGPWAKI 419
486 PHTDGHFHPSPPLMGFGGLKPPQIILKNTVPANPAPESATKFPASFTIYQSTGVSV 545
420 PHTDGHFHPSPPLMGFGGLKPPQIILKNTVPANPAPESATKFPASFTIYQSTGVSV 479
546 IEWELQKNSKRWNPPEVQYTSNVAKSANVDFVNNGLYTEPRPIGRYLTIRPL 599
480 IEWELQKNSKRWNPPEVQYTSNVAKSANVDFVNNGLYTEPRPIGRYLTIRPL 533

RESULT 14

AAM51510 ID AAM51510 standard; protein, 533 AA.

XX AAM51510;

XX 02-JAN-2002 (first entry)

XX Adeno-associated virus VP3 capsid protein.

XX Adeno-associated virus; AAV; capsid; virus-like particle;

KM nuclear localisation signal; VP3.

XX Adeno associated virus.

XX JP2001169777-A.

XX 26-JUN-2001.

XX 30-JUL-1999; 99UP-0249140.

XX 30-JUL-1999; 99UP-0249140.

XX (HANA/) HANA H.

XX WPI; 2001-599854/68.

XX New virus-like particles from VP3 capsid protein of adeno-associated

XX virus, comprise a peptide containing a nucleus-shifting signal

XX connected to its N-terminal -

XX Disclosure; Page 17-19; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a
CC virus-like particle-forming protein, and to a peptide containing a
CC nucleus-shifting signal at its N-terminus that can form virus-like
CC particles by shifting to the nucleus of the animal cell in which it
CC is expressed. The method is used for forming virus-like particles
CC from the VP3 capsid protein of adeno-associated virus (AAV).

SQ Sequence 533 AA;

Query Match 76.2%; Score 2477.5; DB 22; Length 533;

Best Local Similarity 83.3%; Pred. No. 5e-192; Mismatches 52; Indels 1; Gaps 1;

Matches 445; Conservative 36; Mismatches 52; Indels 1; Gaps 1;

66 MASGCGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTYNNHLYKQISS 125
1 MATSGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTYNNHLYKQISS 60
126 ASTGASDNHNYFGYSTPWGTFDNRFHCHSPRDMORLNNNGFPRKLNFLKFNIOVK 185
61 QS-GASNDNHYFGYSTPWGTFDNRFHCHSPRDMORLNNNGFPRKLNFLKFNIOVK 119
186 EVTTNDGVTITANNLSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMVPOGYLTIN 245
120 EVTQNDGVTITANNLSTVQVFSDEYOLPYVLGSAHOGCLPPPADVFMVPOGYLTIN 179
246 NGGQAVGRSSFFCLEYFPSSQMLRTGNNFTSYTFEEVFFHSSVAHSQSLDRLNPLIDQY 305
180 NGGQAVGRSSFFCLEYFPSSQMLRTGNNFTSYTFEEVFFHSSVAHSQSLDRLNPLIDQY 239
306 LYYINRTQNGSGAONKDLFSRGSPPAGMSVQPKMLPGPCYQORVSKTKTDNNNSNFT 365
240 LYYISRTNTPSGITTSRLOFSQAGASDIRDSRNLPGPCYQORVSKTKTDNNNSNFT 299
366 WTGASKTNLNGRESIINPGTAMASHKDEDEKFFPMGSGVMI FGKESAGSNALDNVITD 425
300 WTGATKTHLNGRSLVNPGRPMASHKDEDEKFFPMGSGVMI FGKESAGSNALDNVITD 359
426 EEEIKATNPVATERFGTVAVNFPSSSTDPATGVHAGMALPGVWMDRDVYLQGPWAKI 485
360 EEEIRTTNPVATERFGTVAVNFPSSSTDPATGVHAGMALPGVWMDRDVYLQGPWAKI 419
486 PHTDGHFHPSPPLMGFGGLKPPQIILKNTVPANPAPESATKFPASFTIYQSTGVSV 545
420 PHTDGHFHPSPPLMGFGGLKPPQIILKNTVPANPAPESATKFPASFTIYQSTGVSV 479
546 IEWELQKNSKRWNPPEVQYTSNVAKSANVDFVNNGLYTEPRPIGRYLTIRPL 599
480 IEWELQKNSKRWNPPEVQYTSNVAKSANVDFVNNGLYTEPRPIGRYLTIRPL 533

RESULT 15

AAM46313

ID AAW46313 standard; Protein; 598 AA.
 AC AAW46313;
 DT 28-AUG-1998 (first entry)
 XX
 DE AAV4 VP2 coat protein.
 XX
 KM AAV; AAV-4; VP2; coat protein; vector; gene transfer;
 gene delivery; cancer; gene therapy.
 OS Adeno associated virus 4.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 68 /note= "encoded by GGC"
 FT Misc-difference 114 /note= "encoded by CTN"
 FT
 XX WC09811244-A2.
 PN
 XX 19-MAR-1998.
 PD
 XX 11-SEP-1997; 97WO-US16266.
 PF
 XX 11-SEP-1996; 96US-0025934.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA
 XX Chlorini JA, Kotin RM, Safer B;
 PI WPI; 1998-207403/18.
 XX N-PSDB; AAV21657.
 DR
 XX Adeno-associated virus 4 based vectors - used to transduce erythroid
 PT progenitor cells for treatment of e.g. cancer and other genetic
 PT diseases
 XX
 PS Claim 36; Page 64-66; 80pp; English.
 XX
 CC This polypeptide comprises the adeno associated virus 4 (AAV4)
 CC VP2 coat protein that is encoded by a nucleotide sequence (see
 CC AAV21657) of the AAV4 genome (see also AAV21648). The invention also
 CC provides AAV4 VP1 (see AAW46308) and VP3 (see AAW46314). The invention
 CC provides recombinant vectors and viral particles based on AAV4 that
 CC may be useful for transducing erythroid progenitor cells for the
 CC treatment of e.g. cancer and genetic diseases which can be corrected
 CC by bone marrow transplants using matched donors. A claimed method
 CC of delivering a nucleic acid to a subject (including a subject with
 CC antibodies to AAV2) comprises administering to a cell from the
 CC subject an AAV4 particle comprising the nucleic acid inserted
 CC between a pair of AAV inverted terminal repeats (see AAV21651 and
 CC AAV21659), and returning the cell to the subject. An antibody
 CC against AAV4 VP2 can be used to determine the suitability of an
 CC AAV4 vector for administration to a subject. AAV4 VP2 can also be
 CC used to detect the presence of AAV4-specific antibodies in a sample.
 XX
 SQ Sequence 598 AA;
 Query Match 56.3%; Score 1830.5; DB 19; Length 598;
 Best Local Similarity 57.8%; Pred. No. 1.5e-139;
 Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;
 QY 1 TAPGKKRPVQSPQEPSSGIGTGOAPAKKRLNF-GQTGDSSEVPDPOPLGEPPTPA 59
 1 TAPGKKRPVQSPQEPSSGIGTGOAPAKKRLNF-GQTGDSSEVPDPOPLGEPPTPA 59
 DB 1 TAPGKKRPVQSPQEPSSGIGTGOAPAKKRLNF-GQTGDSSEVPDPOPLGEPPTPA 59
 1 TAPGKKRPVQSPQEPSSGIGTGOAPAKKRLNF-GQTGDSSEVPDPOPLGEPPTPA 59
 QY 60 AVGPTWASGGAGPMANNGSGADGVNAGSGMWHCDSTWMLGDRVITTSRTMALPTYNHL 119
 60 AVGPTWASGGAGPMANNGSGADGVNAGSGMWHCDSTWMLGDRVITTSRTMALPTYNHL 119
 DB 55 MSDDSENRPAAGCAVEGGGAGDGVNAGSGMWHCDSTWSEGHVTTTSTRTWLPTYNHL 114
 55 MSDDSENRPAAGCAVEGGGAGDGVNAGSGMWHCDSTWSEGHVTTTSTRTWLPTYNHL 114
 QY 120 YKOISSASTGASNDNHFGYSTPWGYFDNFNRFHCFSPRDMORLNNMGFRPRKLNFKL 179
 120 YKOISSASTGASNDNHFGYSTPWGYFDNFNRFHCFSPRDMORLNNMGFRPRKLNFKL 179

DB 115 YKILGE-----SLQSNNTYNGSETPWGYFDNFRFHCHFSPRDMORLNNMGFRPRKLNFKL 170
 QY 180 ENIQVEKVTNDGVTTIANNTSTVOVPSDSEYQLPYVLSAAGCLPPRPADVMTIPQY 239
 180 ENIQVEKVTNDGVTTIANNTSTVOVPSDSEYQLPYVLSAAGCLPPRPADVMTIPQY 239
 DB 171 ENIQVEKVTNSGETVANNLTSTVOIFADSSYELPYVMDAGQESLPPEPNDVFWVPQY 230
 171 ENIQVEKVTNSGETVANNLTSTVOIFADSSYELPYVMDAGQESLPPEPNDVFWVPQY 230
 QY 240 GY---LTNNGSQAVRRSSFFCYLEFPSSQMLRTGNFTFSEYEEVPHSSYAHQSGLDR 296
 240 GY---LTNNGSQAVRRSSFFCYLEFPSSQMLRTGNFTFSEYEEVPHSSYAHQSGLDR 296
 DB 231 GYGLVATGNTSQOQTRMNFYCLEFPSSQMLRTGNFTFSEYEEVPHSSYAHQSGLDR 290
 231 GYGLVATGNTSQOQTRMNFYCLEFPSSQMLRTGNFTFSEYEEVPHSSYAHQSGLDR 290
 QY 297 LAMPPLIDQYLYYNRTQN---QSSAQNKDILFSRGSFAGMSVQPRANLPGPCYRQORY 352
 297 LAMPPLIDQYLYYNRTQN---QSSAQNKDILFSRGSFAGMSVQPRANLPGPCYRQORY 352
 DB 291 LAMPPLIDQYLYYNRTQN---QSSAQNKDILFSRGSFAGMSVQPRANLPGPCYRQORY 347
 291 LAMPPLIDQYLYYNRTQN---QSSAQNKDILFSRGSFAGMSVQPRANLPGPCYRQORY 347
 QY 353 SKTKTNNNSNFWTQAS---KY---NNGRESIINPGTAMASHDDEKFFPMGQWI 405
 353 SKTKTNNNSNFWTQAS---KY---NNGRESIINPGTAMASHDDEKFFPMGQWI 405
 DB 348 SKTA--NQYKIPATGSDSLIKYETHSTLDGRMSALTGPMPATGAPASK--FSNSQLIF 404
 348 SKTA--NQYKIPATGSDSLIKYETHSTLDGRMSALTGPMPATGAPASK--FSNSQLIF 404
 QY 406 FGKESAGASNTALDNWITDEEERIKATNPVATERPGTVAVNFQSSSTDPAIGVHMGAL 465
 406 FGKESAGASNTALDNWITDEEERIKATNPVATERPGTVAVNFQSSSTDPAIGVHMGAL 465
 DB 405 AGPKONGNTATVPGTILFTSEBELAATNATDIDMGNLPGQDOSNMLPTVDRLTALGAV 464
 405 AGPKONGNTATVPGTILFTSEBELAATNATDIDMGNLPGQDOSNMLPTVDRLTALGAV 464
 QY 466 PGWVQDRDVLVLOGPIWAKIPHTDGHFHPSPLMGFGKLNPPQILIKNTPVAPNPAEF 525
 466 PGWVQDRDVLVLOGPIWAKIPHTDGHFHPSPLMGFGKLNPPQILIKNTPVAPNPAEF 525
 DB 465 PGWVQDRDVLVLOGPIWAKIPHTDGHFHPSPLMGFGKLNPPQILIKNTPVAPNPAEF 524
 465 PGWVQDRDVLVLOGPIWAKIPHTDGHFHPSPLMGFGKLNPPQILIKNTPVAPNPAEF 524
 QY 526 SATKFPATITQYSTGVSEIEMELQKENSKRNPVQVTSNRYAKSANDFTDNNGLYT 565
 526 SATKFPATITQYSTGVSEIEMELQKENSKRNPVQVTSNRYAKSANDFTDNNGLYT 565
 DB 525 SATKFPATITQYSTGVSEIEMELQKENSKRNPVQVTSNRYAKSANDFTDNNGLYT 564
 525 SATKFPATITQYSTGVSEIEMELQKENSKRNPVQVTSNRYAKSANDFTDNNGLYT 564
 QY 586 EPRPIGTRYLTRPL 599
 586 EPRPIGTRYLTRPL 599
 DB 585 EPRPIGTRYLTRPL 598
 585 EPRPIGTRYLTRPL 598

Search completed: July 17, 2003, 18:31:20
 Job time : 43.664 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 14.7426 Seconds
(without alignments)
1195.466 Million cell updates/sec

Title: US-09-807-802A-15

Perfect score: 3251
Sequence: 1 TAPGKRPEVQSPQEPDSSS.....NGLYTERPRIGTRVTRPL 599

Scoring table: BLOSUM62

Gapop,10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1830.5	56.3	598	4 US-09-532-594B-16	Sequence 16, Appl1
2	1830.5	56.3	734	4 US-09-532-594B-4	Sequence 4, Appl1
3	1690.5	52.0	543	4 US-09-532-594B-18	Sequence 18, Appl1
4	479.5	14.7	543	4 US-08-856-841-22	Sequence 22, Appl1
5	326	10.0	415	4 US-08-856-841-20	Sequence 20, Appl1
6	325	10.0	466	4 US-08-856-841-19	Sequence 19, Appl1
7	325	10.0	500	4 US-08-856-841-16	Sequence 16, Appl1
8	325	10.0	501	4 US-08-856-841-18	Sequence 18, Appl1
9	318	9.8	395	4 US-08-856-841-13	Sequence 13, Appl1
10	310	9.5	264	4 US-08-856-841-14	Sequence 14, Appl1
11	303	9.3	398	4 US-08-856-841-21	Sequence 21, Appl1
12	216.5	6.7	379	6 5223424-13	Patent No. 5223424
13	192	5.9	584	4 US-09-022-949-2	Sequence 2, Appl1
14	178.5	5.5	387	4 US-08-856-841-17	Sequence 17, Appl1
15	113.5	3.5	655	1 US-08-469-202-27	Sequence 27, Appl1
16	113.5	3.5	655	2 US-08-484-434C-34	Sequence 34, Appl1
17	112.5	3.5	3060	2 US-08-487-826B-14	Sequence 14, Appl1
18	110.5	3.4	624	3 US-08-947-965-78	Sequence 78, Appl1
19	110.5	3.4	655	2 US-08-469-202-28	Sequence 28, Appl1
20	110.5	3.4	655	2 US-08-484-434C-35	Sequence 35, Appl1
21	109	3.4	1186	1 US-08-485-568A-4	Sequence 4, Appl1
22	109	3.4	1186	1 US-08-357-698-6	Sequence 6, Appl1
23	109	3.4	1186	2 US-08-580-554A-4	Sequence 4, Appl1
24	109	3.4	1186	2 US-09-184-223-4	Sequence 4, Appl1
25	109	3.4	1186	2 US-09-184-223-4	Sequence 4, Appl1
26	106.5	3.3	1848	5 PCT-US93-12682-6	Sequence 6, Appl1
27	106.5	3.3	1848	5 PCT-US95-10661A-6	Sequence 6, Appl1

28	104.5	3.2	717	4 US-09-626-589-1	Sequence 1, Appl1
29	104	3.2	1013	4 US-09-415-522-8	Sequence 8, Appl1
30	102	3.1	1082	4 US-09-134-001C-3159	Sequence 3159, Ap
31	101	3.1	1651	4 US-09-540-245A-18	Sequence 18, Appl1
32	100	3.1	847	4 US-09-373-157-4	Sequence 4, Appl1
33	99	3.0	824	4 US-09-626-589-3	Sequence 3, Appl1
34	99	3.0	997	1 US-08-232-540-1	Sequence 1, Appl1
35	99	3.0	997	1 US-08-428-949-1	Sequence 1, Appl1
36	99	3.0	997	1 US-08-428-948A-1	Sequence 1, Appl1
37	99	3.0	997	2 US-08-428-948-1	Sequence 1, Appl1
38	99	3.0	997	5 PCT-US95-04656-1	Sequence 1, Appl1
39	99	3.0	998	1 US-08-233-008A-6	Sequence 6, Appl1
40	99	3.0	1021	1 US-08-233-008A-2	Sequence 2, Appl1
41	98.5	3.0	644	1 US-08-206-176-2	Sequence 2, Appl1
42	98.5	3.0	1000	4 US-08-891-640-3	Sequence 3, Appl1
43	98.5	3.0	1036	4 US-08-891-640-3	Sequence 3, Appl1
44	98.5	3.0	1091	3 US-08-633-768A-2	Sequence 2, Appl1
45	98	3.0	1096	4 US-09-415-946-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1

US-09-532-594B-16

Sequence 16, Application US/09532594B

Patent No. 6468524

GENERAL INFORMATION:

APPLICANT: Chordin, John A.

APPLICANT: Kottin, Robert M.

APPLICANT: Safer, Brian

APPLICANT: Davidson, Beverly

TITLE OF INVENTION: AA4 VECTORS AND USES THEREOF

CURRENT APPLICATION NUMBER: US/09/532,594B

CURRENT FILING DATE: 2000-03-22

NUMBER OF SEQ ID NOS: 22

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16

LENGTH: 598

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =

NAME/KEY: misc.feature

OTHER INFORMATION: AA4 capsid protein VP2

US-09-532-594B-16

Query Match

Best Local Similarity 57.8%; Score 1830.5; DB 4; Length 598;

Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;

QY	1	TAPGKRPEVQSPQEPDSSSGIGKTQOPAKKRLNF-CQTGSESVDDPOPLGEPATPA	59
DB	1	TAPGKRPEVQSPQEPDSSSGIGKTQOPAKKRLNF-VEDEGADGPPESTSG-----A	54
QY	60	ANGPTTMAAGGAPMADNNEGADGVGNASGNHCDSTWLADRYITTTSTRTALPTNNHL	119
DB	55	MSDSEMRRAAGAAVEGQAGDGVGNASGNHCDSTWEGHVTITSTRTALPTNNHL	114
QY	120	YKQISSASGASNDNHPYSPFWGVPFDRNRFCHFSPPDMQRLINNNGFRKRLNFXL	179
DB	115	YKRLGE-----SLOSTNTNGSTFWGVPFDRNRFCHFSPPDMQRLINNNGFRKRLNFXL	170
QY	180	FNIOVEKVTNDGVTIANNLSTVQVPSDEXQPLVLSAHOGLPPFPADVFMIPQY	239
DB	171	FNIOVEKVTTSNGETTANNLSTVQIFADSSYELPYWDAGGEGSLPPFPADVFMIPQY	230
QY	240	GY---LTLNNGSQAVERSSFYCLEYPPSQMLRTGNNFTSYTBEVPHSSVYHSGSLDR	296
DB	231	GYCGLVGTGNTSQOQOTDRNAFYCLEYPPSQMLRTGNNFTSYTBEKVPFHSVYHSGSLDR	290

QY 297 LAMP.LIDQYLYLNRQN-----OSGSAQNDL.FSRGSPAGMSVQPPKWLPGPCYQORV 352
DB 291 LAMP.LIDQYLYLNRQSTTTGTTLNAGTATN---FTKLRPTNFNFKNWLPGBSIIQOQGF 347
QY 353 SKRTDNNNSNFTWTGAS---KY-----NLNGRESIINPGTAMASHKDEDEKFFPMGCVMI 405
DB 348 SKTA--NQNKIPATGSDSLIKETHTSLDGRMSALTPGPMATAGADSK-FSNSQLIF 404
QY 406 FKGESAGASTALDNWMTDEEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHANGAL 465
DB 405 AGPKQNGNTATVPGLTIFTESEELAAATNATDITMGMNLPGGDOSNSNLPVDRLLTALGAV 464
QY 466 PGWMOODRDVYLOGPIWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPAEF 525
DB 465 PGWMOODRDVYLOGPIWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPAEF 524
QY 526 SATKFASTFOYSTGOVSVEIEMELQENSKRNPNPEVQYTSNYSKANSVDFVNNGLYT 585
DB 525 SSTPVSFTFOYSTGOVSVDIWEIQERSKRNPNPEVQYTSNYSQNSLIMAPDAAGKYT 584
QY 586 EPRPIGTRYLTP 599
DB 585 EPRPIGTRYLTHL 598

RESULT 2
US-09-532-594B-4

Sequence 4, Application US/09532594B
Patent No. 6468524
GENERAL INFORMATION:
APPLICANT: Chorini, John A.
APPLICANT: Kotlin, Robert M.
APPLICANT: Safer, Brian
APPLICANT: Davidson, Beverly
TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
FILE REFERENCE: 14014.0252U2
CURRENT APPLICATION NUMBER: US/09/532.594B
CURRENT FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 734
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
NAME/KEY: misc feature
OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

Query Match 56.3%; Score 1830.5; DB 4; Length 734; .
Best Local Similarity 57.8%; Pred. No. 3.6e-154; Indels 31; Gaps 10;
Matches 355; Conservative 74; Mismatches 154;
QY 1 TAPKPKPVEOSPPSPSSGIGKTGOOPAKKRLNF--GQTSISEVDPDQPLGPPATPA 59
DB 137 TAPKPKPLIESPPQPSSTIGIKKQKQPAKKLVFEDETAAGGPPBGISG-----A 190
QY 60 AVGETTASGGAGMADNMGADGVGNASGNWCHDSITWLGDEVITTSRTWALPTYNHNL 119
DB 191 MSDSEKRAAAGAAVGGAGDAGVGNASGDMHCDSTWSEGHVTTSTRVWLPYNNHNL 250
QY 120 YKQISSASTASNNHFGYSTPWGYDPFNRFFCHFSRDMQRLINNMMGRPRKLPK 179
DB 251 YKRLGE---SLQNTNGSTFWGYDPFNRFFCHFSRDMQRLINNMMGRPRKLPK 306
QY 180 FNIQKQVTTNDGVTIANNLSTVOVSDSEYQLPYVLSAHOGLPPPADVFMIPOY 239
DB 307 FNIQKQVTTNSNGTITVANNLITSTVOIFADSSYELPYVMDAKGQGSLLPPFNDVFMPOY 366
QY 240 GY---LTLNNGSQAVGSSFYCLEYFPQOMLRTGNNTFTSYTPEEVPHSSYAHQSILDR 296

DB 367 GYCVLVTGNTSQOQTDNRNAFYCLEYFPQOMLRTGNNTFTSYTPEEVPHSSYAHQSILDR 426
QY 297 LAMP.LIDQYLYLNRQN-----OSGSAQNDL.FSRGSPAGMSVQPPKWLPGPCYQORV 352
DB 427 LAMP.LIDQYLYLNRQSTTTGTTLNAGTATN---FTKLRPTNFNFKNWLPGBSIIQOQGF 483
QY 353 SKRTDNNNSNFTWTGAS---KY-----NLNGRESIINPGTAMASHKDEDEKFFPMGCVMI 405
DB 484 SKTA--NQNKIPATGSDSLIKETHTSLDGRMSALTTPGPMATAGADSK-FSNSQLIF 540
QY 406 FKGESAGASTALDNWMTDEEBEIKATNPVATERFGTVAVNFQSSSTDPAATGVHANGAL 465
DB 541 AGPKQNGNTATVPGLTIFTESEELAAATNATDITMGMNLPGGDOSNSNLPVDRLLTALGAV 600
QY 466 PGWMOODRDVYLOGPIWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPAEF 525
DB 601 PGWMOODRDVYLOGPIWAKIPHTDGHFHPSPLMGFGKLPKPPQILIKNTVPANPAEF 600
QY 526 SATKFASTFOYSTGOVSVEIEMELQENSKRNPNPEVQYTSNYSKANSVDFVNNGLYT 585
DB 601 SSTPVSFTFOYSTGOVSVDIWEIQERSKRNPNPEVQYTSNYSQNSLIMAPDAAGKYT 720
QY 586 EPRPIGTRYLTP 599
DB 721 EPRPIGTRYLTHL 734

RESULT 3
US-09-532-594B-18

Sequence 18, Application US/09532594B
Patent No. 6468524
GENERAL INFORMATION:
APPLICANT: Chorini, John A.
APPLICANT: Kotlin, Robert M.
APPLICANT: Safer, Brian
APPLICANT: Davidson, Beverly
TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
FILE REFERENCE: 14014.0252U2
CURRENT APPLICATION NUMBER: US/09/532.594B
CURRENT FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 18
LENGTH: 544
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
NAME/KEY: misc feature
OTHER INFORMATION: AAV4 capsid protein VP3
US-09-532-594B-18

Query Match 52.0%; Score 1690.5; DB 4; Length 544;
Best Local Similarity 59.4%; Pred. No. 6.4e-142;
Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;
QY 67 ASGGAPMADNMGAGGVGNASGNWCHDSITWLGDEVITTSRTWALPTYNHNLKYQISSA 126
DB 9 AAAGGAIV--EGGQAGAGVGNASGDMHCDSTWSEGHVTTSTRVWLPYNNHLYKRLGE-- 66
QY 127 STGASNDNHFGYSTPWGYDPFNRFFCHFSRDMQRLINNMMGRPRKLPK 186
DB 67 ---SLQNTNGSTFWGYDPFNRFFCHFSRDMQRLINNMMGRPRKLPK 123
QY 187 VTTNDGVTIANNLSTVOVSDSEYQLPYVLSAHOGLPPPADVFMIPOY---LT 243
DB 124 VTTNSGRTVANNLITSTVOIFADSSYELPYVMDAGQGSLLPPFNDVFMPOYCGGLVT 183
QY 244 LNNGSQAVGSSFYCLEYFPQOMLRTGNNTFTSYTPEEVPHSSYAHQSILDR 303
DB 184 GNTSQOQTDNRNAFYCLEYFPQOMLRTGNNTFTSYTPEEVPHSSYAHQSILDR 243

QY 304 QYLYINRTON----QSGSAONKDLFSPGSPAGMSVOPKXNWLPGPCROORVSKTKTDN 359
 Db 244 QYLMGLSTTGTTLNLTATN---FTKLRPTNFSNFKXNWLPGPSIKQOGSKTA--N 298
 QY 360 NNSNFTTGS--KY---NLNRESIINPTGMASHKODEDEKFPMSGVMIFGKESAG 412
 Db 299 QNKKIPATGSDSLKXETHSTLDRMSALTPGPMATAGPADSK-FSNSOLIFAGPKONG 357
 QY 413 ASMTALDNWMTTDEBEIKATNPVATERFGTVAVNFOSSTDPAATGDVHAGMLPGMWOD 472
 Db 358 NTATVETITLFTSEELATNATDITDWMGNLPGSDQSNLPTVDRLLTALGAVFGWON 417
 QY 473 RDVYLOGPIWAKIPHTDGHFHPSPLMGFGGLKNPPOILIKNTPVPANPPAESATKFA 532
 Db 418 RDIYOGPIWAKIPHTDGHFHPSPLMGFGGLKNPPOILIKNTPVPANPPATTSSTPVNS 477
 QY 533 PITQYSTGOVSVEIEMELQKENSRRANPEVOYTSNVAKANVDFTVNNGLYTEPRIGT 552
 Db 478 PITQYSTGOVSVOIDWEIOKERSKRNPEVOFTSNYQONSLMWPADAGKYTEPRAIQT 537
 QY 593 RYLTRPL 599
 Db 538 RYLTHHL 544

RESULT 4

US-08-856-841-22
 ; Sequence 22, Application US/08856841
 ; Patent No. 6274307
 ; GENERAL INFORMATION:
 ; APPLICANT: ERWIN SOUTSCHEK
 ; APPLICANT: MANFRED MOTZ
 ; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
 ; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
 ; STREET: 99 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10016
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" FLOPPY DISC
 ; COMPUTER: AT&T - IBM COMPATIBLE
 ; OPERATING SYSTEM: MS-DOS Version 6.2
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/856, 841
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/214, 658
 ; FILING DATE: 16-MARCH-1994
 ; APPLICATION NUMBER: US 07/917, 096
 ; FILING DATE: 4-AUGUST-1992
 ; APPLICATION NUMBER: PCT/DE91/00106
 ; FILING DATE: 8-FEBRUARY-1991
 ; APPLICATION NUMBER: DE40038262
 ; FILING DATE: 8-FEBRUARY-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROBINSON, WILLIAM R.
 ; REGISTRATION NUMBER: 27,224
 ; REFERENCE/DOCKET NUMBER: LKR-9222-A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 697-3355
 ; TELEFAX: (212) 557-5635
 ; TELEX: NONE
 ; INFORMATION FOR SEQ ID NO: 22:
 ; LENGTH: 543
 ; TYPE: AMINO ACID
 ; TOPOLOGY: LINEAR

MOLECULE TYPE:
 DESCRIPTION: PEPTIDE
 HYPOTHEtical: N/A
 ANTI-SENSE: N/A
 FRAGMENT TYPE: INTERNAL
 ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
 IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
 POSITION IN GENOME: N/A
 FEATURE:
 NAME/KEY: N/A
 LOCATION: N/A
 IDENTIFICATION METHOD: amino acid analysis and
 IDENTIFICATION METHOD: mass spectrometry
 OTHER INFORMATION:
 PUBLICATION INFORMATION:
 AUTHORS: COSSART, Y.E.
 AUTHORS: FIELD, A.M.
 AUTHORS: CANT, B.
 AUTHORS: WIDDONS, D.
 TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
 JOURNAL: LANCET
 VOLUME: 1
 ISSUE:
 PAGES: 72 - 73
 DATE: 1975
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO: 22:
 US-08-856-841-22

Query Match 14.7%; Score 479.5; DB 4; Length 543;

Best Local Similarity 27.1%; Pred. No. 3,4e-34;

Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16;

QY 75 ADNNEGADGVGNAS--GNMHCSTWIGDRVITTSRTWALPTFNNHLKYKQISSASTGASN 132
 Db 16 AEAATGAGGGGSGSVSMSEGAATFANSVTCFSGQFLIPDPREHHYKFSABAASCHN 75
 QY 133 D-----NHFGYSTPMGTFDFNFRCHSPSPDQRLINNMGFPKRLNFKLFINI 182
 Db 76 ASGEAKVCTIIPIMGYSTPMRYLDENALNLFPSPLFQHLIENYGIAPDALTVTISEI 135
 QY 183 QVKEAT--INDGYTTANNLTSTVOYFSDSEYQLPVVLSAHQCLPPRPADVFMTPOYG 240
 Db 136 AVKDVTDKXGGGV-QVTDSTTGRCLMVLVHEKYKYPVVLQGGDTLAPBLPIWVYFPPOYA 194
 QY 241 YLTANN-GSOAVG-----RSSFYCLEYFSPQWLRTGNNFTFSYTFEEVFPFSSYAH 290
 Db 195 YLTGCVNTQGISGDSKLAESASATVLEHSSFOLGTGTRASNSYKFPVPPENLBGC 254
 QY 291 SGLDLRLNPLIDQVLYINRTONOGSAONKDLFSPGSPAGMSVOPKXNWLPGPCYRQ 350
 Db 255 SQHFYEMVNP---YGSRLGVPTLGGDPKFRSL-----THEDHAIQPNFMFGPLVNSV 306
 QY 351 RVSKTIDNNNSFTTGTASKTNLNGRESIINPG-TAMASHKODEDEKFPMSGVMIFGK 409
 Db 307 STYEGDSSNTGAKALTLGTSTGSTRSLRPGVOPDYHHMDTDKVTGAINAISHGOT 366
 QY 410 SAGASNTALDNV-----MTDEBEIKATNPVATERFGTVAVNFOSSTDPAATGDVH 462
 Db 367 TYG--NAEDKEVQGVGRFNEKEKLOLOGLNHTY-----FNNKGTQOYTDQIR-R 416
 QY 463 GALPGWODRDVYLOGPIWAKIPHTDGHFHPSP-LPMGFGGLKNPPOILIKNTPVPAN 521
 Db 417 PLWVGSVNNRRALHYESQMSKIPNLDSEFKTOFALGGMGLHQPPOI-----465
 QY 522 PAEFSATKRASFTQYSTGOVSVEIEMEL-QKENSRRANPE 561
 Db 466 -----FLKQYAVGIMTVTMTFKLGPRKATGRANPO 495

RESULT 5
US-08-856-841-20
Sequence 20, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 415
TYPE: AMINO ACID
MOLECULE TYPE: LINEAR
DESCRIPTION: PEPTIDE
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOMAS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73

DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 20:
US-08-856-841-20
Query Match 10.0%; Score 326; DB 4; Length 415;
Best Local Similarity 26.0%; Pred. No. 1e-20;
Matches 101; Conservative 66; Mismatches 180; Indels 42; Gaps 13;
QY 193 VTTIANNLSTVOVDFDSEYLPVLSAHOGLPPPADVEMIPQYGYTLNN-GSOAV 251
Db 1 MTMTISLHACMLV--DHEKYPYVLGGQDDTLAELPIWVPPQYALTVGDVNTQGI 58
QY 252 G-----RSSFCLEIYPPSQMLTGNFTSYTTEEPFHSSVAHSGLRLNPLI 302
Db 59 SCDSKLASESAFYVLEHSSFOLLGTGTAASNYKPPVPPNLEGCSQHFYEMYNPL- 117
QY 303 DOYLYVLRNQNSGSAONDDLFSRSGPAGMSVOPKMWLPFGCYRQQRVSKTKTDNNNS 362
Db 118 --YGRUGVPTDTGGDPKFRSL-----THEDHALQPNFMGPVLSVSTKEDSSNTGA 170
QY 363 NPTWTGASKYXNLNGRESIINPG-TAMASHKDDKDFPMGVMIFGESAGASNTALDNV 421
Db 171 GKALTGSTGTSQNTRLSLAPGPVSPQYHHMDTDKVTGINAISHQDTYG--NAEDKE 227
QY 422 -----MTDEEIKATNVATERGTYAVNFGSSSTDPAQDVIAMGALPGMWQDRD 474
Db 228 YQGVGHPKPKQLQGLNHTY-----FPNKGTOQYTOIE-RPLMVGSVMNRR 280
QY 475 VYLQGPWAKIPHTDGHFHS-PLMGFGKPNPOLLINPTVPANPPEFATYFASF 533
Db 281 LHYESQLMSKIPVLDSDFKQFALGQWGLHQPPIFLK--ILPESGPIGKISGIT 338
QY 534 ITQYSTGVSEIEMEL-QKNSKRNPE 561
Db 339 LVQYAVGIMVTMTFKLGPRAKATGRNMPQ 367
RESULT 6
US-08-856-841-19
Sequence 19, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991

APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 486
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 19:
US-08-856-841-19

Query Match 10.0%; Score 325; DB 4; Length 486;
Best Local Similarity 32.4%; Pred. No. 1.6e-20;
Matches 90; Conservative 33; Mismatches 121; Indels 34; Gaps 7;
QY 51 LGPPPTPAVGTPTTASGGGAPMADNNGADYGNASG-----WHCDSTWLGDRYITT 105
DB 213 LPVPVYVNASSEKYPSTMTSVNSA-----EASTGAGGGGSPVXSMSEGTFSANSTCT 266
QY 106 STRTWALPTNNHLYKOISSASTGASND-----NHVPGYSPMGYEDNRRHCHP 155
DB 267 FSRQFLIPYPEHHKYFSPASSCHNASKAKVCTISIPWGSIFWRILDNALNLF 326
QY 156 SPRDWRLINNNWGFRRKRLNFKLNFIOVEVT--TNDGYTTANMLTSTVOVPSDEYO 213
DB 327 SPLFQHLINENYSTADALVTIISLAIVADVTDKTGGV-QVTDATGRICMLVHREYK 385
QY 214 LPVYLSAHGCGCLPPPADVFMTPOYGYTLNAN-GSAVAV-----RSSYCLEYFP 263
DB 386 YPVYLGQGGDTLAPELPIWVYPPQYAVYLVGVDVNTQGISDASKLASESAYVLEHSS 445
QY 264 SOMLRGNNFTSYTFEEVFPFSSVYAHSSQSLDRMLNPL 301
DB 446 FOLLGTGATMTSKFPVPEPENIEGSCQHFYEMNPL 483

RESULT 7

US-08-856-841-16
Sequence 16, Application us/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFENER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 500
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975

DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION INFORMATION:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
JOURNAL: GENE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
US-08-856-841-16

Query Match 10.0%; Score 325; DB 4; Length 500;
Best Local Similarity 32.2%; Pred. No. 1.7e-20;
Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;
QY 51 LGEPPTPAVNGPTTMASGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITTSRTW 110
DB 221 LPEVPAYNASKPYMTSVNSAE-ASTGAGGGGSSNSVKSMSGATPSANSVCTCFEROF 279
QY 111 ALPTNNHLYKQISSASTGASND-----NHFGSTPMGYPFDRFCHFSRDM 160
DB 280 LIPDPPEHHYVFPFPAASCSNAGSKKAKVCTISPIWGISTPMMYLDPMALNLFESLDF 339
QY 161 QRLNNNGFRRKLNFKLNFNIQVEVT--TNDGVTITANN,ISTVOVSDSEYQLPYVL 218
DB 340 QHLIENYGSIAPDALTVTISEIAVADVTDKTGGGV-QVTDSYTGRLCMLVDHEYKYPYVL 398
QY 219 GSAHOGCLPPRPAPVFMIPQGYLTINN-GSQANG-----RSSFYCLEYFPSSQMLR 268
DB 399 GQGQDITAPLPIWYFPQYAVLTIVGQVNTQGISGSKLASSESAFYVLEHSSQDLG 458
QY 269 TGNNTFSYTFEEVYFHSYAHQSILDLNPL 301
DB 459 TGGTASMSYKRPYPVPENLBCSGHFYEMNPL 491

RESULT 8
US-08-856-841-18
Sequence 18, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE

CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 18:
US-08-856-841-18

Query Match 10.0%; Score 325; DB 4; Length 501;
Best Local Similarity 32.2%; Pred. No. 1.7e-20;
Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;
QY 51 LGEPPTPAVNGPTTMASGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITTSRTW 110

Db 222 LPEVPAYNAERKYPMTSUNSAB-ASTGAGGGSNSVKNMSEGAFTSANSVCTFGRQF 280
Qy 111 ALPTNNHLYKQJSSASTGASND-----NHFGYSPWPGYPPENRFCHFSRDM 160
Db 281 LIIPDPFHNYVFPSPASSCHNASKKAKVCTTISPINGSTPMYLDPMNLNFPSPLEF 340
Qy 161 QRLNNMNGFRPKLNFKNLQVKEVT--TNDGVTTIANNLSTVOVFDSEYQLPYVL 218
Db 341 QHLENGSTAPDMLTYTISINAKVDYTDKIGGV-QVDTSTGRCLMVDHEKXYVL 399
Qy 219 GSAHOGCLPPPADVEMIPQYGYTLNN-GSQAVG-----RSSFYCLEYFPSPQMR 268
Db 400 GCGQDTLAPELPIWVYPPQYAYLTGVDVNTQGISGSKLASSESAFYVLEHSSFQLLG 459
Qy 269 TGNNTFTSYTEEBYFPHSSYAHQSGLRLNPL 301
Db 460 TGGTASMSYKFPVPVPPENLEGSCQHFYEMYNPL 492

RESULT 9
US-08-856-841-13
Sequence 13, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LTR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 395
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE

ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDMANS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 13:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 13:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
JOURNAL: GLUTATHIONE S. TRANSFERASE
VOLUME:
ISSUE:
PAGES: 67
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 13:
US-08-856-841-13

Query Match 9.8%; Score 318; DB 4; Length 395;
Best Local Similarity 25.4%; Pred. No. 4,8e-20;
Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;

Qy 193 VTTIANNLTSTVOVFDSEYQLPYVLSAHOGLPPFPADVEMIPQYGYTLNN-GSQAV 251
Db 1 MMTITPSLHACMLV--DHEKXYVVGQGDPLAPBLPIWVFPQYAYLTGVDVNTQGI 58
Qy 252 G-----RSSFYCLEYFPSPQMR-----TNNFTSYTEEBYFPHSSYAHQSGLRLNPLI 302
Db 59 SGDSKSLASEESAFAVYVLEHSSFOLLGTGTASMSYKFPVPVPPENLEGSCQHFYEMYNPL- 117
Qy 303 DOVLYYINANTONGSGAOKNDLFSRSGPAGMSVQPKMLPGCYNQAVSKTKTNNNS 362
Db 118 --YGSRLGVDPDTIGSPKFRSL---THEDHAIQONFMGGLVNSVSTKESDSSNTGA 170
Qy 363 NFTWTGASKYKNIINGRESIINPG-TAMASHKDEDEKFPMSGVMTFKESAGASNTALDNV 421

Db 171 GKALTGSTGTSQNTLRISLRPGVSPYHMDTKYTGAINAISHGQTTG---NAEDKE 227
Qy 422 -----MTDEEELKATNPVATERFGTVAVNFGSSSTDPTGVDHMGALPGWMDRD 474
Db 228 YQOGVGPRPNKEQOLKQLOGNMHTY-----FPKGTQOYTDIE-RPLWVGSVNRRRA 280
Qy 475 VYLOGPIAKIPIHTDGHFHPs-PLMGFGGLKNPPQILKILTPVANPFAFSATKFSASF 533
Db 281 LHYESQMSKIPNDSDSKTQPAALGGWGLHOPPOI-----F 318
Qy 534 ITQYSTGVSVYEIMEL-QKENSRRMDE 561
Db 319 LKQYAVGIMTVMTFKLGPRKATGWRNPO 347
RESULT 10
US-08-856-841-14
Sequence 14, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LRR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 264
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
POSITION IN GENOME: N/A
FEATURE:

NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDONS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
US-08-856-841-14
Query Match 9.5%; Score 310; DB 4; Length 264;
Best Local Similarity 31.3%; Pred. No. 1.3e-19; Mismatches 104; Indels 34; Gaps 6;
Matches 78; Conservative 33;
Qy 57 TPAVGPPTWASGGA PMADNNEGADVGNASGNWCHDSITWLGDRVITSTRTMALPTYN 116
Db 5 TPGLAAASGTGAGG-----GSNSVKSMESEGAIFFSANSTCTCSRQFLIPYDP 53
Qy 117 NHYKOISASGTASND-----NHYEGYSTWGYDFENRFCHSPSPRMORLIIN 166
Db 54 EHHYKVFSPASSCHNASGKAQCTISPIMGYSTWRKYDFNALNLFSPLEFOHLIEN 113
Qy 167 NMGFRPKRLNFKLNFYQVEVT--TNDGVTIANNLITVQVSDSEYOLFVYLGSAHQG 224
Db 114 YGSIAPDALVTITSELAIVKQVDTKGTGGV-QVTDSTTGRICMLVDHEKYKPYVLGGQODT 172
Qy 225 CLPFPADVPMIOYGLTLIN-GSOAVG-----RSSFYCLEYFPQMLRTGNFT 274
Db 173 LABELPFWYFPFOYALVYTGVDVNTQISGDSKLAASEBAFVLEHSSFQLLTGTAS 232
Qy 275 FSYTFEEVP 283

Db 233 MSYKFPVP 241

RESULT 11
US-08-856-841-21
Sequence 21, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUSCHER
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 398
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I

ISSUE: 72 - 73
PAGES: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 21:
US-08-856-841-21

Query Match 9.3%; Score 303; DB 4; Length 398;
Best Local Similarity 25.4%; Pred. No. 1,1e-18;
Matches 96; Conservative 48; Mismatches 168; Indels 66; Gaps 12;

QY 209 DSEYOLPYVIGASHQGLPEPPADVPMIQGYLTNN-GSQAVG-----RSSFC 258
Db 4 DHEYKTPYVIGQGDLTAPLPIWTFPPQYALVIGDVNTGIGSDSKLASESAFYV 63
QY 259 LKYPQMLRTGNNFTFSYTFEEVPHSSVAHSQSLRLNPLIDYLYIANRTONQGS 318
Db 64 LEHSSFQLGTGTATMSTYFFPPVPPENLEGCQHYEMYNPL---YGSRLGVPTLIGD 120
QY 319 AQNKDLFSRGSFAGSVQPKWLPGPCYRQOVSTKTDNNNSNFTWGASKYNNGRE 378
Db 121 PKFRSL-----THEDAIQPNFMPLVNSVSTKSGDSNTGAGLITGLSTGTSQNT 175
QY 379 SIINPG-TMAASHKDDKDFEPMGWMIFGKESAGASNTALDNV-----MITDEEK 430
Db 176 ISLRPGVSQPIYHMDTKVTGINAISGQTYG---NADREYQGVGRFPNEKQLK 232
QY 431 ATNPVATERGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPIMAKIPHTDG 490
Db 233 QLOGLMHTY-----FPNKGTQYTDQIE-RLIMGVSNNRRLAHYESGLMSEKINLD 285
QY 491 HPHPS-PLMGFGLKNPPQILIKNTVPANPPAEFATKFAFTQYSTGQVSEIEME 549
Db 286 SFKTOPAALGWMGLHPPQI-----FKYHKVQGLEVLNQWE 323
QY 550 LQENSK-----RWNP 560
Db 324 LPPNMPWELQHLNMP 341

RESULT 12
5223424-13
Patent No. 5223424
APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
RICHARD D.
TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
ACID SEQUENCE
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 27-JUL-1988
FILING DATE: 27-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 78,519
FILING DATE: 27-JUL-1987
APPLICATION NUMBER: 933,107
FILING DATE: 20-NOV-1986
APPLICATION NUMBER: 902,887
FILING DATE: 02-SEP-1986
APPLICATION NUMBER: 887,140
FILING DATE: 17-JUL-1986
APPLICATION NUMBER: 823,102
FILING DATE: 27-JAN-1986
APPLICATION NUMBER: 773,430
FILING DATE: 06-SEP-1985
SEQ ID NO: 13:
LENGTH: 579
5223424-13
Query Match 6.7%; Score 216.5; DB 6; Length 579;
Best Local Similarity 22.6%; Pred. No. 9.3e-11;

Matches 143; Conservative 84; Mismatches 250; Indels 157; Gaps 33;

Qy 42 SESVPDDPQLPEPRATPAVGPPTMAAGGAPMADNNEGADGVGNAGNMH--CDSTWLG 99
 Db 2 SENVEQHNPINAGTEL-SATNESGGGGGGG---GGRGAGGVSTGTFNNQTEFFQYL 56
 Qy 100 D---RVITTSRTWALPTYNHLYKQIS--SASTGAS---NDNHFGYSTPMGYPFNR 150
 Db 57 EGVLRITAHASRLHLHNHEHETRYKRIHVNSSEGVAGQWQDADHQMVTPLMLIDRNA 116
 Qy 151 FHCPSPPDQRLINNMGFRPKRLNPLFNIQVKEVT---TNDGVTIANNLSTVOVF 207
 Db 117 WGVFNPADQOLISNNMTEINLVSEFQEIFNVVLTXTITESATSPSKYNNDLTASLMAVA 176
 Qy 208 SDSFOLPYVLGSAHQGLPPRPADVEMIPQYGLV-----TLNNGSOAVGRS--- 254
 Db 177 LDTNNLTPTPAAPRSETLGFYPMLPKPTQRYLSCIRNLNPTTYGSSQOITDSIQ 236
 Qy 255 -----SFYCLE-YFPQOMLRTGNNF-TFSYTFEVPFHSSVAHSOSLDRMLNPLIDQYL 306
 Db 237 GLHSDIMXYTIENAVPIHLRTGDEFSIGIHFDTKPL-KLTHSMQTNR----- 284
 Qy 307 YILNRTONQSSAONKDLFRRGSPAGMSVOP-----KMLPGPCYRQ---QVSKTK 356
 Db 285 -----SLGLPPKVLTEPTTEGDHPGLPGANTRKQYHQTINNSY 324
 Qy 357 TD-----NNNSFTWTGASKXNGLGRESIINP-GTMAHSHKDEKFPFMSG 402
 Db 325 TEATALRPAQVGYTPYNNFDYSNGGPF-----LTPYVFTADTOYIDE---PWGA 372
 Qy 403 VMIFKESAGASANTALDNMTDEBEIKATNP-----VATERGTVA-VNFOSSS--- 451
 Db 373 IRFTMGVGHMLTTS-----SQELERYTFNPQSKGRAPKQGNQOAPLMLNTNNGTL 426
 Qy 452 --TPPAG--DVHANGALP-----GMWQDDVYLOQPIYAKIHTD--GHFSPMLMG 500
 Db 427 LPSDPIGGKSNKHFMTNLTNYGPLTALNNTAPVFPNGQIMDKELDTDKPRILH--VYAP 483
 Qy 501 FGLG-NPPOLLINNTPVANPAPAFSA--TKPASFITYQSTGVQSVIEMLKENSKR 557
 Db 484 FVCKNNPFGQLFVKIAP---NLTDFFNADSPQOPIIID-SNFMWKGITLTATAKRSNM 539
 Qy 558 WNPVQYTSNYAKSANDFTVDNNGLYTEPRPIG 591
 Db 540 WNPQOHT-----TAEINRKVI-PTNIG 562

RESULT 13
 ; Sequence 2, Application US/09022949
 ; Patent No. 6187759
 ; GENERAL INFORMATION:
 ; APPLICANT: Tarpey, Ian
 ; APPLICANT: Greenwood, Neil
 ; TITLE OF INVENTION: Canine parvovirus DNA vaccination
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Akzo No. 6187759el Patent Dept.
 ; STREET: 1300 Piccard Drive, Suite 206
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: US
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30(EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/022,949
 ; FILING DATE: 11-FEB-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:

NAME: Gormley, Mary E.
 REGISTRATION NUMBER: 34,409
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-948-7400
 TELEFAX: 301-948-9751
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 584 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-022-949-2

Query Match 5.9%; Score 192; DB 4; Length 584;
 Best Local Similarity 19.9%; Pred. No. 1.4e-08;
 Matches 128; Conservative 100; Mismatches 232; Indels 184; Gaps 32;

Qy 49 QPLGEPPA--TPAVGPTTMAAGGAPMADNNEGADGVGNAGNMH--CDSTWLG 99
 Db 7 QPDGQPAVANERATGSGNGSGGGG-----GGSGGVGISTGFNNQTEFFKLENGWV- 59
 Qy 100 DRVITTSRTWALPTYNHLYKQI-----SSASTGASNDNHFGYSTPMGYPFNR 152
 Db 60 -EITANSRLVHLMPESENYRRVVNNLDKTAINGMALDTHAQIVTPWLSVDANAWG 118
 Qy 153 CHSPRPMORLINNMGFRPKRLNPLFNIQVKEVT---TNDGVTIANNLSTVOVFS 209
 Db 119 VNRFGDMQILVMTMSLHVSFEQEIFNVVLTXTSESATOPPTKYNNDLTASLMAVD 178
 Qy 210 SEYQLPYVLGSAHQGL-----PPPPA-----DVEMIPOYGLTLNNGSOAVGRS- 254
 Db 179 SNNMTPTPAMRSETLGFYPMKPTIPTWRYFFQMDRTLIPSH-----TGTSQPTNI 232
 Qy 255 -----SFYCLE-YFPQOMLRTGNNF-TFSYTFEVPFHSSVAHSOSLDRMLNPLID 303
 Db 233 YHGTDPDDVQFYTIENSVPVHLRTGDEFATGTFEDCKP-CRLTHWQTNALG--LP 288
 Qy 304 QVLYLNRTONOS-----GSAONK-----DLFRRGSPAGMSVOPKMLPG 344
 Db 289 PFLNSLPQSGGNTFGYIGVQDKRQGYTMGNTNITTEITIRPAVGS-----A 340
 Qy 345 PCYRQQRVSK-----TKTDNNNSFTWTGASKY---NLGRESIINPGT--- 385
 Db 341 PYSFBASTGPPKPTPIAAGRGAGQDENQAA---DGDPRYARGHQGQYTTTGETPER 397
 Qy 386 -AMASHDDDKFPPSGWMI PKESAGASNTALD-NVMTTDEBEIKATNPVATERGTIV 443
 Db 398 FYTIAHQDT-----GRYPEGDIQININFLPVTNNVLLPTDPG---GKT 440
 Qy 444 AVNFQSSSTDPATGDVHANGALPGMWQDDVYLOQPIYAKIHTDGHFSPSLMGFGL 503
 Db 441 GINY--TINITYGPLTALNVP-----PVYFNQOIMDKEDTD-----L 478
 Qy 504 K-----NPPOLLINNTPVAN--PPAFSATKFASTFYQSTGVQSVIE 547
 Db 479 KPLRHVAPFVCCNCGQLFVKAAPLITREYDPPDASAMSR-----IVTSDPMWKGKIV 534
 Qy 548 WELQKSKRWNPVQYTSNYAKSANDFTVDNNGLYTERPIG 591
 Db 535 FKAKLRASHTNPIQWMSIN-----IDNQFNV-PSNIG 567

RESULT 14
 ; US-08-856-841-17
 ; Sequence 17, Application US/08856841
 ; Patent No. 6274307
 ; GENERAL INFORMATION:
 ; APPLICANT: ERWIN SOUTSCHEK
 ; APPLICANT: MANFRED MOLT
 ; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
 ; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
 ; NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 387
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 17:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMEROOK, J.
TITLE: MOLECULAR CLONING

JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 15:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 17:
US-08-856-841-17
Query Match 5.5%; Score 178.5; DB 4; Length 387;
Best Local Similarity 31.6%; Pred. No. 1,2e-07;
Matches 48; Conservative 17; Mismatches 76; Indels 11; Gaps 2;
QY 51 LGEPATPAVAGPTWASGGGAPMADNNEGADGVGNASGWNHCDSTWLGDRVITSTRW 110
Db 221 LPEVPAYNASSEKPTSMTVNSAE-ASTGAGGGSSNVKSMBSGATFSANSTCTSRQP 279
QY 111 ALPTNNHLYKOISSASTGASND-----NHVFGYSTPWGYPDFNRFHCHFSPPDM 160
Db 280 LIPYDEHHYKVPSPASSCHNAGSKAKVCTTSPIMGYSTPRYLDPNALNLFSPDLER 339
QY 161 QRLINNMGRPRKALNFKLENIOKVEVTNDG 192
Db 340 QHLIENYGSIAPDALTVTISEIAVKDVTDTKG 371
RESULT 15
US-08-469-202-27
Sequence 27, Application US/08469202
Patent No. 5750875
GENERAL INFORMATION:
APPLICANT: STALKER, DAVID
TITLE OF INVENTION: GLYCOSYL BIOSYNTHETIC ENZYMES IN PLANTS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALGENE, INC.
STREET: 1920 FIFTH STREET
CITY: DAVIS
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1 (a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,202
FILING DATE: 6-JUNE-95
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/016,881
FILING DATE: 11-FEB-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Lassen

Job time : 16.7426 secs

REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
NAME: Carl J. Schwedler
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: CGNE 93-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 916-753-6313
TELEFAX: 916-753-1510
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 655 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-469-202-27

Query Match 3.5%; Score 113.5; DB 1; Length 655;
Best Local Similarity 18.4%; Pred. No. 0.16; Mismatches 257; Indels 225; Gaps 35;
Matches 129; Conservative 89;

QY 12 SPOEDSSGIGKTGQOPAKKRLNFQGTGSESVDPDPIGEPATPAVGPPTMASGGG 71
DB 54 SDGDPNNAGFNSATYDNNLKKTG--GDLRGLINKLPYLK-----SLGVTSIWI--- 102
QY 72 APNADNNEGADGVNAS--GWHCDSTMLGDRV-----ITSTRMALPTNNHLYKOIS 124
DB 103 TPIPIDNNNTDAAGNTGYGWRDYRIDHFGNLDKELTSLMSPDYNMRLVLDYA 162
QY 125 SASGTASNDNH-----FGYSTPWGYFDENRFGHFSPRDMQRLINN 168
DB 163 PNHNSANDENFGALYRDGVITTDYPTDVANTGM-----YHNGGVTNW-----ND 209
QY 169 GFRPRRLNFKLENIQVKEVTNDVTTIANNLTSTVOVPSSEYOLPYVLGSAH--QGCL 226
DB 210 FFQYK--NHNLFNS---DLNQSNTDVOYLIDGSKFWIDAGVDAIRIDAIKMDKSF 263
QY 227 PRPADVFMIPQYGYLTINNGSAVGRSSFYCL--EYFPGMLRT-----GN----- 271
DB 264 QKWTSDI---YDY-----SKSIGREGFFPGSEWFGASANTTGVGNADIDYANTSGS 312
QY 272 ---NFTSFYFEEVFPFSSVYAHSSQSLDRLMNPILIDQYLYL----- 309
DB 313 ALDIFGRDITLRLVGRS---GNTMKTILNSYLIRQTVFTSDMQVFMNDHMANIGT 369
QY 310 ---NRT---QNGSSAQNKDILLFSR--GSPAGMSVOPFNNLPGBCTYRQORVSKTKT 357
DB 370 ALRGNATTFPGNNNETGSGSEAFQKRIIDLGLVATMTVR--GIPAIYGTETHYAANFT 426
QY 358 DNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPM----- 400
DB 427 SN---SPQGVSDPYN--REKM--PGF-----DTESEAPSIITKLGLDKSSPAIONG 472
QY 401 ---SGWIFGKESA-----GASNTA-LDNWMTD-----EEIYKATN 433
DB 473 TYTELWVNDLIVERRSGNDIYVALNRCGANTINVINAVPGVPSLIGNNSVAVAN 532
QY 434 PVAT-ERRGIVAVNFQSSSTDPATGDVHAM--GALPGWQDRDYYLOGPIWAKIPHTDG 490
DB 533 KOATLTLMQNEAVVIRSQSDAENPTQISINFACNNGYTIISGQSVYIIGNI----- 583
QY 491 HFHSPMLMGFGGLNPPQIILIKNTPVANPAPERSATKFASTITQYSTGQVSE----- 545
DB 584 ---POLGGWDLTK---AVKISP-----TQYFQWSASLELBSDL 615
QY 546 -IEWELOKENGKRWNPVEVQYTSNTAKSANVDFTVDNNGLY 584
DB 616 NVEKCVARNETNPANVEMQSGANNQNSNDTQTTNGSF 555

Search completed: July 17, 2003, 18:36:38

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 28.8443 Seconds
(without alignments)
2466.245 Million cell updates/sec

Title: US-09-807-802A-15
Perfect score: 3251
Sequence: 1 TABGKRPEVQSPQEPDSSS.....NNGLYTEPRPIGTRVLRPL 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues
Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764.5	85.0	735	US-10-038-972A-13	Sequence 13, Appl
2	2764.5	85.0	735	US-10-293-478-1	Sequence 1, Appl
3	2759.5	84.9	538	US-10-038-972A-14	Sequence 14, Appl
4	2481.5	76.3	533	US-10-038-972A-15	Sequence 15, Appl
5	772	23.7	756	US-10-205-942-4	Sequence 4, Appl
6	118	3.6	2344	US-09-815-242-12713	Sequence 12713, A
7	116.5	3.6	5795	US-09-815-242-12610	Sequence 12610, A
8	108	3.3	1463	US-09-971-536-69	Sequence 69, Appl
9	106.5	3.3	1848	US-09-839-996-6	Sequence 6, Appl
10	106.5	3.3	1848	US-10-080-505-6	Sequence 6, Appl
11	106.5	3.3	2364	US-10-156-761-7834	Sequence 7834, Ap
12	106	3.3	420	US-10-062-254-254	Sequence 254, App
13	104.5	3.2	717	US-10-046-583A-1	Sequence 1, Appl
14	103.5	3.2	1016	US-09-815-242-5845	Sequence 5845, Ap
15	103.5	3.2	1265	US-10-198-070-69	Sequence 69, Appl
16	102	3.1	3712	US-10-108-605-103	Sequence 103, App

17	101.5	3.1	439	US-09-836-353A-145	Sequence 145, App
18	101.5	3.1	439	US-09-984-130-145	Sequence 145, App
19	101.5	3.1	1232	US-09-801-574-46	Sequence 46, Appl
20	101	3.1	676	US-09-801-368-302	Sequence 302, App
21	101	3.1	1411	US-10-080-505-17	Sequence 17, Appl
22	100.5	3.1	1714	US-10-098-916A-11	Sequence 11, Appl
23	100.5	3.1	4019	US-09-738-973-425	Sequence 425, App
24	100.5	3.1	4019	US-09-854-133-425	Sequence 425, App
25	100.5	3.1	4019	US-10-144-649A-425	Sequence 425, App
26	100	3.1	847	US-10-112-527-4	Sequence 4, Appl
27	99	3.0	623	US-10-108-605-125	Sequence 125, App
28	99	3.0	623	US-10-108-605-129	Sequence 129, App
29	99	3.0	824	US-10-046-583A-3	Sequence 3, Appl
30	99	3.0	1228	US-09-917-384-1	Sequence 1, Appl
31	99	3.0	1228	US-09-917-383-1	Sequence 1, Appl
32	98.5	3.0	515	US-10-097-340-212	Sequence 212, App
33	98.5	3.0	515	US-10-171-311-156	Sequence 156, App
34	98.5	3.0	620	US-09-931-009A-1	Sequence 1, Appl
35	98.5	3.0	644	US-09-919-039-121	Sequence 121, App
36	98.5	3.0	1000	US-10-055-412B-30	Sequence 30, Appl
37	98.5	3.0	1036	US-09-842-256-3	Sequence 3, Appl
38	98.5	3.0	1091	US-09-280-197-2	Sequence 2, Appl
39	98.5	3.0	1091	US-09-423-126-2	Sequence 2, Appl
40	98.5	3.0	2478	US-09-815-242-5816	Sequence 5816, Ap
41	98.5	3.0	2478	US-09-815-242-15967	Sequence 12967, A
42	98	3.0	425	US-10-156-761-12331	Sequence 12311, A
43	98	3.0	1770	US-09-841-132-444	Sequence 444, App
44	97.5	3.0	685	US-09-745-763-11	Sequence 11, Appl
45	97	3.0	659	US-10-046-583A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-038-972A-13
; Sequence 13, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36960US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 735
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VPI caaspd protien
US-10-038-972A-13

Query Match	85.0%; Score 2764.5; DB 15; Length 735;
Best Local Similarity	Pred. No. 4.3e-235;
Matches	496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;
QY	1 TABGKRPEVQSPQEPDSSSGIGKTQQAARKLNFGQTGDSVDPQPLGEPATPAA 60
DB	138 TABGKRPEVHSEVPESSGTSKAGQAPARKLNFGQTGADSVDPQPLGEPAPSG 197
QY	61 VGTTHASGGAGMADNNEADVGNAAGNHCDSITWLDGRTVTTSTRTALPTNNHLY 120
DB	198 LGNTMAATGSGAMANNBAGADVNGSGNWHCDSTWMDGRVTTSTRTALPTNNHLY 257
QY	121 KOISASTGASNDNHFGYSTPWGDFPNRFCHFSPPDMQRLINNMGRRPKLNFKLF 180
DB	258 KOISASGASNDNHFGYSTPWGDFPNRFCHFSPPDMQRLINNMGRRPKLNFKLF 316
QY	181 NIQKVEVTTNDGVTIANNTISTVQVPSDESYQLPYVLSAGHGGCLPPFADVFMIPOYG 240
DB	317 NIQKVEVTTNDGVTIANNTISTVQVFTDSEYQLPYVLSAGHGGCLPPFADVFMIPOYG 376

QY 241 YLTINNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 300
DB 377 YLTINNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 436
QY 301 LIDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 360
DB 437 LIDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 496
QY 361 NSFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420
DB 497 NSESWTGATKYHLNGRDSLVPNPGAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 556
QY 421 VMTDEEIKATNPVATERFCTVAVNFQSSSTDPAATGVDVAMGALPGMWQDRDVLQGP 480
DB 557 VMTDEEIKATNPVATERFCTVAVNFQSSSTDPAATGVDVAMGALPGMWQDRDVLQGP 616
QY 481 IMAKIPHTDGHFHPSPLMGFGGLKXNPPOILLIKNTVPANPAPAFSAATKFASTTQYSTG 540
DB 617 IMAKIPHTDGHFHPSPLMGFGGLKXNPPOILLIKNTVPANPAPAFSAATKFASTTQYSTG 676
QY 541 QVSVEIEMELQKNSKRNPEVOYTSNYAKASANDFTVDNNGLYTERPRIGTRILRPL 599
DB 677 QVSVEIEMELQKNSKRNPEVOYTSNYAKASANDFTVDNNGLYTERPRIGTRILRPL 735

RESULT 2

US-10-293-478-1
; Sequence 1, Application US/10293478
; Publication No. US20030078411A1
; GENERAL INFORMATION:
; APPLICANT: PATEL, SATI D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/10/293, 478
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/321, 589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 735
; TYPE: PRN
; ORGANISM: Adeno-associated virus
US-10-293-478-1

Query Match 85.0%; Score 2764.5; DB 15; Length 735;
Best Local Similarity 82.8%; Pred. No. 4.3e-235;
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPKKRPVEQSPQEPDSSSGIGTKGOAPAKRLNFGQTDGSESVPPDQPLGEPAPAA 60
DB 138 TAPKKRPVEQSPQEPDSSSGIGTKGOAPAKRLNFGQTDGSESVPPDQPLGEPAPAA 197
QY 61 VGPTTASGGAPADNNEGADGVGNASGNHCDSTWLGDRVITTSRTMALPTYNNHLY 120
DB 198 LGTMTATGSGAPADNNEGADGVGNASGNHCDSTWLGDRVITTSRTMALPTYNNHLY 257
QY 121 KQISSASTGASNDNHYFGYSTPWGYPFDPNRFCHFSPPDWQRLINNMGFPKRLNFKLF 180
DB 258 KQISSAS-GASNDNHYFGYSTPWGYPFDPNRFCHFSPPDWQRLINNMGFPKRLNFKLF 316
QY 181 NIQKEVYQNDGTTIANNLSTVOVFSDEYQLPYVLGSAHQCLPFPADVEMIPQYX 240
DB 317 NIQKEVYQNDGTTIANNLSTVOVFSDEYQLPYVLGSAHQCLPFPADVEMIPQYX 376
QY 241 YLTINNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 300
DB 377 YLTINNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 436
QY 301 LIDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 360

DB 437 LIDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 496
QY 361 NSFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420
DB 497 NSESWTGATKYHLNGRDSLVPNPGAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 556
QY 421 VMTDEEIKATNPVATERFCTVAVNFQSSSTDPAATGVDVAMGALPGMWQDRDVLQGP 480
DB 557 VMTDEEIKATNPVATERFCTVAVNFQSSSTDPAATGVDVAMGALPGMWQDRDVLQGP 616
QY 481 IMAKIPHTDGHFHPSPLMGFGGLKXNPPOILLIKNTVPANPAPAFSAATKFASTTQYSTG 540
DB 617 IMAKIPHTDGHFHPSPLMGFGGLKXNPPOILLIKNTVPANPAPAFSAATKFASTTQYSTG 676
QY 541 QVSVEIEMELQKNSKRNPEVOYTSNYAKASANDFTVDNNGLYTERPRIGTRILRPL 599
DB 677 QVSVEIEMELQKNSKRNPEVOYTSNYAKASANDFTVDNNGLYTERPRIGTRILRPL 735

RESULT 3

US-10-038-972A-14
; Sequence 14, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/3696US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 14
; LENGTH: 598
; TYPE: PRN
; ORGANISM: adeno-associated virus 2 VP2 capsid protein
US-10-038-972A-14

Query Match 84.9%; Score 2759.5; DB 15; Length 598;
Best Local Similarity 82.8%; Pred. No. 8.7e-235;
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APTKKRPVEQSPQEPDSSSGIGTKGOAPAKRLNFGQTDGSESVPPDQPLGEPAPAA 61
DB 2 APTKKRPVEQSPQEPDSSSGIGTKGOAPAKRLNFGQTDGSESVPPDQPLGEPAPAA 121
QY 62 GPTTASGGAPADNNEGADGVGNASGNHCDSTWLGDRVITTSRTMALPTYNNHLY 121
DB 62 GPTTASGGAPADNNEGADGVGNASGNHCDSTWLGDRVITTSRTMALPTYNNHLY 121
QY 122 QISSASTGASNDNHYFGYSTPWGYPFDPNRFCHFSPPDWQRLINNMGFPKRLNFKLF 181
DB 122 QISSAS-GASNDNHYFGYSTPWGYPFDPNRFCHFSPPDWQRLINNMGFPKRLNFKLF 181
QY 182 IYQKEVYQNDGTTIANNLSTVOVFSDEYQLPYVLGSAHQCLPFPADVEMIPQYX 241
DB 181 IYQKEVYQNDGTTIANNLSTVOVFSDEYQLPYVLGSAHQCLPFPADVEMIPQYX 240
QY 242 LITNNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 301
DB 241 LITNNGSQAAGRSSFYCLEYFSPQMLRTGNNFTSYFEEVPHSSVAHSQSLDRLMNP 300
QY 302 IDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 361
DB 301 IDQYLYLSTRTNTPSGTTTQSRLOFSQAGASDIRDOSRMLPGPCYRQOQVSKTSADNN 360
QY 362 SNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 421
DB 361 SEVSWTGATKYHLNGRDSLVPNPGAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420
QY 422 MTDDEEIKATNPVATERFCTVAVNFQSSSTDPAATGVDVAMGALPGMWQDRDVLQGP 481

Db 421 MITDEERIRTNPVATQYGSVSTNLQRGNQAATAADVTQVLPGMWODRDVYLQGP 480
QY 482 WAKIPHTDGHFHPSPGLMGFGGLKMPPOILLIKNTVPANPAEBSATKPSFITYSTGQ 541
Db 481 WAKIPHTDGHFHPSPGLMGFGGLKMPPOILLIKNTVPANPAEBSATKPSFITYSTGQ 540
QY 542 VSVEIEMELQKENSCKRNPEVOYTSNYSKASANDPTVDNNGLYTEPRPIGRYLTBP 599
Db 541 VSVEIEMELQKENSCKRNPEVOYTSNYSKASANDPTVDNNGLYTEPRPIGRYLTBP 598

RESULT 4

US-10-038-972A-15
; Sequence 15, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/3696US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 533
; TYPE: PRP
; ORGANISM: adeno-associated virus 2 VP3 capsid protein
US-10-038-972A-15

Query Match 76.3%; Score 2481.5; DB 15; Length 533;
Best local Similarity 83.3%; Pred. No. 2.5e-210; Gaps 1;

Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 66 MASGGAGPAMADNNEGADGVNAGSNHCHDSTWLDGRTVITSTRTWALPTYNNHLYKQISS 125
Db 1 MATGSGAPMADNNEGADGVNAGSNHCHDSTWLDGRTVITSTRTWALPTYNNHLYKQISS 60
QY 126 ASFGASDNHNYFGSTWGWGFDPRFRHCHSPRMQGLINNMGFRKRLNFKLNFQVK 185
Db 61 QS-GASDNHNYFGSTWGWGFDPRFRHCHSPRMQGLINNMGFRKRLNFKLNFQVK 119
QY 186 EYTTNDGVTIANNLTSTVOVFSDEYQLPYVLSAHOGLPPPADVPMIPOXYGLTN 245
Db 120 EYVONDGTTIANNLTSTVOVFTDSEYQLPYVLSAHOGLPPPADVPMIPOXYGLTN 179
QY 246 NGSOAVGRSSFYCLEYPSQMLRTGNNFTFSYTEEVPFHSYAHSGSLDRLNMLIDQY 305
Db 180 NGSOAVGRSSFYCLEYPSQMLRTGNNFTFSYTEEVPFHSYAHSGSLDRLNMLIDQY 239
QY 306 LYLTLNRQNSGSAONKDLPSRGSFAGMSVOPKMLPGPCYRQORYSKTKTNNNSFT 365
Db 240 LYLTLNRQNSGSAONKDLPSRGSFAGMSVOPKMLPGPCYRQORYSKTKTNNNSFT 299
QY 366 WTGASCKNLNLRGSSIIINPGTAMASHKDDKPPMSGVMI FGKESASATLADNMTD 425
Db 300 WTGATKYLNLGRSOLVNPFGPAMASHKDDKPPMSGVMI FGKESASATLADNMTD 359
QY 426 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMALPGMWODRDVYLQGPWAKI 485
Db 360 EEEIRTNPNVATERFGTVAVNFQSSSTDPATGDVHAMALPGMWODRDVYLQGPWAKI 419
QY 486 PHTDGHFHPSPGLMGFGGLKMPPOILLIKNTVPANPAEBSATKPSFITYSTGQVSE 545
Db 420 PHTDGHFHPSPGLMGFGGLKMPPOILLIKNTVPANPAEBSATKPSFITYSTGQVSE 479
QY 546 IEMELQKENSCKRNPEVOYTSNYSKASANDPTVDNNGLYTEPRPIGRYLTBP 599
Db 480 IEMELQKENSCKRNPEVOYTSNYSKASANDPTVDNNGLYTEPRPIGRYLTBP 533

RESULT 5

US-10-205-942-4
; Sequence 4, Application US/10205942
; Publication No. US20030053930A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053930A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205, 942
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 756
; TYPE: PRP
; ORGANISM: Adeno-associated virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2271)
; OTHER INFORMATION: B19/AAV chimeric capsid coding sequence
US-10-205-942-4

Query Match 23.7%; Score 772; DB 15; Length 756;
Best local Similarity 32.9%; Pred. No. 3.1e-59;
Matches 195; Conservative 87; Mismatches 257; Indels 54; Gaps 16;

QY 1 TABGKRPEVQSPQEPDSSSGIGKTGOQPAKRLNFGQTDSESVDPQGLGPPATPA 60
Db 138 TABGKRPEVQSPQEPDSSSGIGKTGOQPAKRLNFGQTDSESVDPQGLGPPATPA 197
QY 61 VGTWMASSGGAAMADNNEGADGVNAGSNHCHDSTWLDGRTVITSTRTWALPTYNNHLY 120
Db 198 LGTWTIVNSA-AAGTAGGGGNSVKSMSGATFASVCTCSROFLIYDEBNH 256
QY 121 KOISASTGASND-----NHFGYSTPMGYDPRFRHCHSPRMQGLINNMGFR 170
Db 257 KVPSPAASSCHMASGKEAKVCTISPIMGSTPWRKYDFALNLFSPLEOHLENYGSI 316
QY 171 RPRRLNFKLNFQVKEVT--TNDGVTIANNLTSTVOVFSDEYQLPYVLSAHOGLPP 228
Db 317 APDALTVITSEIAYKQVTDKGTGGV-QVTDSTTGRCLMVDHRYKYPVYLGQGDTLAPE 375
QY 229 FPDVPMIPOXYGLTNL-GSOAVG-----RSFYCLEYPSQMLRTGNNFTFSY 278
Db 376 LPIWVYFPQYALTVGDVNTQGISGDSKRLASESAFVLEHSPQLGTGTATMSYK 435
QY 279 FEEVPHSSYAHSGSLDRLNMLIDQYLYLNRQNSGSAONKDLPSRGSFAGMSVOP 338
Db 436 FEPVPEPNLEGCQHYTEMNPPL---YGSRLGVPTDLGDPKRLSL-----THEDAIOP 487
QY 339 KMLPGPCYRQORYSKTKTNNNSFTWTGASCKNLNLRGSSIIINPG-TAMASHKDDK 397
Db 488 QNFMGPPLVNSVSTKGDSNSTGAKALGLSTGTSQNRISLRPQVPSQPYHHMTDXY 547
QY 398 FPMGSMITGKESAGASNTALDNV-----MTDEEIKATNPVATERFGTVAVNFQSS 450
Db 548 VTGINAISHGQTYTG---NAEDKEYQGVGRFNEKEOLKQLOGLMHTY-----FPMK 598
QY 451 STDPATGDVHAMALPGMWODRDVYLQGPWAKIHTDGHFHPSPGLMGFGGLKMPPO 509
Db 599 GTQOYTDQIE-RLMTGSMYNNRRALYEQSLMSKLNLDSEFTQALGKGLHQPPO 657
QY 510 ILIKNTVPANPAEBSATKPSFITYSTGQVSEIEMEL-QKENSCKRNPE 561
Db 658 IFLK--ILPQSGBIGIKSMGITTLIVQAVAMVTMTFKLGRKATGRMNPQ 708

RESULT 6
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12713
LENGTH: 2344
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 3.6%; Score 118; DB 10; Length 2344;
Best Local Similarity 20.4%; Pred. No. 0.89;
Matches 87; Conservative 57; Mismatches 182; Indels 100; Gaps 18;

37 GGTGDSSEVP-DPQPLGEPPTPAVG-----PTMASGGGAPMA 75
326 GNGDGGIGFAPSPGVLGETGNGAVGIGLSNFGKLDYHNTSTNSGAKAKADPSN 385
76 DNEGADGVGNAGMHCDSITWLGDRVITTSRTWALPTNNHLKYQISSASTGASNDNH 135
386 VAGGAFAPFTTDSYVASTYSSSTADNAKLVQFTNN-----TFODPDIN 434
136 YFGYSTPWGDFDFFRCHFSRDMQRLINNNWGFRRPLNFKLENIQVKEVTNDGVTT 195
435 YNG-----DTKVMYKAGQWTNRNI-SDWLAKSGTNNPSL-----SMTASTGATN 480
196 IANNITSTVOVPSDSBYOLPYVLSAHQGLP--FPADVEMIPQYGYLTIANLSQAVGR 253
481 LQOVQFGFEXTESAVTQVRVDTTGKDIIPPKYSGNVQV-----VTIDNOQSALTA 535
254 SSF-----YCLEYFPS-----QMLRTGNNFTSYTEFEEVPHSSVYAHQSGLRLMPL 301
536 KGVNTSYDSYASTYNDINKTKVAKTNAGOSYTFITFDVKAFTYVNGQITLEVGMTNPI 595
302 IDOYLTYLNRTONSGAONKDLFSRGSAPGMSV-QPKMWPGPCYR--QORVSKTKTD 358
596 V-----LTTDNGGIVTNT-----VTGLPSGLSDSATNSILIGPTKIGGSTVTVSTD 645
359 --ANNSNTWTGASKYLNNGRESIIN-----PGTAMASHODDEKFPFMSGVMIFGKES 410
646 QANNKSTTFT-----INVDTTAPVTPTIGDKSS-VFSPISPIINAIQDN 691
411 AGASNT 416
692 SGNAV 697

RESULT 7
US-09-815-242-12610
Sequence 12610, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12610
LENGTH: 5795
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 3.6%; Score 116.5; DB 10; Length 5795;
Best Local Similarity 19.1%; Pred. No. 4.6;
Matches 128; Conservative 86; Mismatches 250; Indels 207; Gaps 35;

40 GDSSEVDPQPLGEPPTPAVGPTTMASGGAPMADNEGADGVGNAGMHCDSITWLG 99
1696 GNGSSADYKLSNGSAIPRA---TITWVGQAPNKDNTRIGEDT-NTANILID-----G 1747
100 DRVITTSRTWALPTNNHLKYQISSASTGASNDNHYFGYSTPWGDFDFFRCH----- 154
1748 ETTPITKTAL-----YKTVS-----SVKPHVEETRGAVFPGVSD 1782
155 -FSPRDMQRLINNNWGFRRPLNFKLENT-----QVKEVTNDGVTTIANLSTVOVS 208
1783 VYDAKQVKEVNDNSWTQMAQRMNFQFTNSYGPBKDVAGSTRIDIRVYDHOQIILIA 1842
209 DSEYQLEPVYVLSA-----HQGLPFPADVEMIPQYGYLTIANLSQSA 250
1843 KVPKDEPRIDGNSVYTKAGLTNQIKINNVLSSSSIKLFPAD--NTP-----LITNTTYG 1896
251 VGRSSFYCL-EYPSQMLRTGNNFTSYTEFEEVPHSSVYAHQSGLRLMPLIDOYLTYL 309
1897 SGNTAVVTVSDALPNGVYKARS-----SITMNNVTYTTQDEHGAIDVTRRESVD----- 1946
310 NRTONSGS-----AQKDLFSRGS-----PAGMSV-----QPKMW- 341
1947 ---SNDSATVYTPOLQATTEGAVFIKGGDGFPGHVEREIQNPFGATVAMHNDPTWK 2003
342 -----LPQPCYRQORVSKTKTDNNNSFTWAGSKYLNNGR 377
2004 NTGANTHTKTAIVVTLPSGGQTRAVVEVPKYVPVANAAPASDVAGQULT-----NGT 2054
378 ESI-----INPGT-----AMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNVIT-- 424

Db 2055 DAINTYTFDNTVNTNGITAAWNRQCPNQQ--QAGVQHLNVDTYVGTIAAKRVPVTYN 2111
Qy 425 -DDEEIKATNPVATERGTAVANFQSSSTPARGDVAMAKA--LP--GMV--WQDRDVL 477
Db 2112 VYGFEPQTSYTTTVG--GTILANGTQ-----ASGYAMONANGLPDTGFTYKNNMATGT 2164
Qy 478 QGPIWAKI--PHT-----DGHFHSPLMGSGFGLKN--PPQILIKNTPVPA--- 519
Db 2165 NDANWMAAMKPNNAKVNAKYNDVIYNQHTATSLPAKFVVKDVQPAKPTVETIAGAITI 2224
Qy 520 NPAPESATKFASTFIQYS-----TGQYSVEIEMELQENSKRNNEVQYTSNVAKSA 572
Db 2225 TPGANQVNVTHAGNVTTTADKLVIKKNQNV---VTIFRNNNTSPWKE-----ASATVA 2277
Qy 573 NVDPFTVDNGL 583
Db 2278 GIAGT--NNGI 2286

RESULT 8

US-09-971-536-69
; Sequence 69, Application US/0971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Hanukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 104362
; CURRENT APPLICATION NUMBER: US/09/971, 536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634, 238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724, 623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

Query Match 3.3%; Score 108; DB 11; Length 1463;
Best Local Similarity 21.4%; Pred. No. 3.4; Indels 138; Gaps 27;
Matches 114; Conservative 67; Mismatches 214

Qy 1 TABGKRPRVQSPQEPDSSGIGKTGOQPAKKILNFQGTDSBSVDPDPLGEPPTPAA 60
Db 85 TSSGSSDITTSQANEKNSASVTSASANSASATNTSBSQT-----FVNBAVSNDAS 134
Qy 61 VGTPTMASGGAGMADNNEAGDVGASGWHGCDSTLGPVITTSRTALPTYNH-- 118
Db 135 SADVSTASEPDAAMADSTV--SVINVQSPFVMDVS--GGRQ--SYAERPMLIINGNNHTI 188
Qy 119 -LYKQISSASTGAS-----NDNHVGYSTPWGYPDFNFHGFSPRDMQRLINNMGF 170
Db 189 DPGKXFFHAPDPTSSQNSFTINDLMWGYG-----W-----WG- 221
Qy 171 RPRRLNFKLENIQVKEVTNDGV--TTIANNLTSTVQVFSDESEYOLPYVLGSAHOGCLPP 228
Db 222 -----PYTIKSGKPDGIDHSVVFNNVTYTGAOLMWGYTKAFIKGNT----- 264

Qy 229 FPADEVMIPOYG-YLTUNGS-----QAVGRSSPYCLEYFPSPQMLRTGNNPFTSYTFEEVY 283
Db 265 -----KIOSVGVSYPPLDOSTTQTGLGNQNFQIStYLEVLPGATYTTGTCINVEY 318
Qy 284 FHSSVA-----HSQSLDRLMNPPLIDQVYYLNRTONQSGAQNKLPSRGS 330
Db 319 DGGSFVVDKATVNLRTDASKSNERGTALID-----TQGGNVEFKD-----GS 363
Qy 331 PAGMSQPKWLP-----GPCYRQRVSKTKTDNNNSNFTW--TGASKYNLNGREST--INPG 384
Db 364 TV-----ILNNKALVKDGFAPLYIEDGGNLTVDKNATVITGATGNIPRIDGTGVNLNEG 420
Qy 385 TAAASHKDEDEKFEFSGWVIFKESAG-----ASNTALD--NVWITDEBEIKATNPATER 439
Db 421 ---SHMTTQNGAPKLGTFINIKGTGPFVWSSGLDLNVGTGTSVNAIN--VANDG 475
Qy 440 FGTVAVNFQSSST-DEATGPDVHAMGALPGWMDR--DVTLOGPIMAKIPTHT 489
Db 476 QLSFAQDATANLITDGTGEAHL-----KVGDDANINIMPKSVLFKITDND 523

RESULT 9

US-09-839-996-6
; Sequence 6, Application US/0983996
; Publication No. US20030009010A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Plent, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839, 996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296, 791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 761-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-09-839-996-6
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 3.3%; Score 106.5; DB 12; Length 1848;
Best Local Similarity 18.7%; Pred. No. 6.4; Indels 237; Gaps 31;
Matches 133; Conservative 79; Mismatches 264

Qy 60 AVGPPTMASGGAGPMA--DNNEGAD-----GVGNASGNWHD----- 94

Db 363 ATGSTITTGGGEPPLSVLDLTDGKDPNHGKSIITLKSGTLTANNHIDQAGGLFFEGDYE 422
Qy 95 -----STWIGDRVITTT-STRTMAL--PTY----- 115
Db 423 VKGTSSTWKGAGVADGKTVTWKVNHPKYDLAKIGKGLVVEGKNGEGLLKVDG 482
Qy 116 -----NNHL--YKQ--ISSASTGASNDNH-----YFGYSTPMGFDFNRFHC 153
Db 483 TVLKQKADANNKQVQSVQVIGSRSTLVANDKQVDPNSITFGFR--GGRDLNLSL 540
Qy 154 HF-----SPRDMQRLINNNMGFRPRKLNFKLNIOVKEVTTNDVTTIANNITSTVOVPSD 209
Db 541 TFDHINIDGAVVNHNM-----TNTSNITIT--TGESLITNPNTITSYINIEAOD 588
Qy 210 SEYQLPVYLSAHQGLPPRPADVPMIPOGYITLNNGSA-----VG 252
Db 589 DDH--PLRIRSI-----PYROLYFNQDNRSYTLTKKASTRSELPONGSESNENMLYMG 640
Qy 253 RSSEYCLEYFPSSQMLRTGNFTFSYTEFEVPHSSVAHSQSLDRMLNPLIDQYLYLNR 312
Db 641 RTSDAARVNWNHINNNRMNGFNCFGEE--ETKATQNGKLVNTPNGSKDQNRFLTGG 697
Qy 313 QNOSGSAQ-NKDLLFSRGSF-----AGMSVQPKWMLPGPCYRQORVSKTKTDNNNSNF 364
Db 698 TNLNGDLNVEKGTFLSGRTPPHARDIAGISTTKD---PHFTENNEVVVEDDWINRNF 753
Qy 365 TWGASRYNNGRESIIN-----PGTAMASHKDEDEKFFPMSCVMIFGESAG----- 412
Db 754 K--ATTMNVYGNASLYSGRNVANITSNTITASNNAQVHIGYKTDVTCVRSDDYGVYTC 810
Qy 413 ---ASNALD-----NMTIDEDEI--KATNPATERFGTVAVNFQSSSTDPATG 457
Db 811 NSNLSEKALSNFNTNLKGNVNLTEHNSFTLGKANTLGTSITQSVNLKENSHPMLTG 870
Qy 458 DVHMGALPGMWQDRDVLVLOGPIWAKIPHTDGHFHPSPMLGSGGLKNPPOILLKNTPV 517
Db 871 NSNV-----NQLNLNTHGIH-----LNQNDANKVTYVTTLT 902
Qy 518 PANPAPFASATKFSF-----ITQYSTGVASVEIELOKENSKRANPEVQVTSNY 568
Db 903 VNSLSGNGSYFYWDFTNNKSNKVNVNKSATGNFTLOV--ADKTGEPNHNETLTFDASN 959
Qy 569 AKSANDVFTVDN-----NGLYTEPRP-----IGTRYLTRP 598
Db 960 ATRNNLEVTLANGSVDRGAMKYKLRNVNGRDLYNPEVEKNGQVDTTNTITP 1012

RESULT 10
US-10-080-505-6
; Sequence 6, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080.505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1848
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-6

Query Match 3.3%; Score 106.5; DB 15; Length 1848;
Best Local Similarity 18.7%; Pred. No. 6.4;
Matches 133; Conservative 79; Mismatches 264; Indels 237; Gaps 31;

Qy 60 AVGPETWASGAGAPMA-DNNEGAD-----GVGNASGMWHD----- 94
Db 363 ATGSTITTGGGEPPLSVLDLTDGKDPNHGKSIITLKSGTLTANNHIDQAGGLFFEGDYE 422
Qy 95 -----STWIGDRVITTT-STRTMAL--PTY----- 115
Db 423 VKGTSSTWKGAGVADGKTVTWKVNHPKYDLAKIGKGLVVEGKNGEGLLKVDG 482
Qy 116 -----NNHL--YKQ--ISSASTGASNDNH-----YFGYSTPMGFDFNRFHC 153
Db 483 TVLKQKADANNKQVQSVQVIGSRSTLVANDKQVDPNSITFGFR--GGRDLNLSL 540
Qy 154 HF-----SPRDMQRLINNNMGFRPRKLNFKLNIOVKEVTTNDVTTIANNITSTVOVPSD 209
Db 541 TFDHINIDGAVVNHNM-----TNTSNITIT--TGESLITNPNTITSYINIEAOD 588
Qy 210 SEYQLPVYLSAHQGLPPRPADVPMIPOGYITLNNGSA-----VG 252
Db 589 DDH--PLRIRSI-----PYROLYFNQDNRSYTLTKKASTRSELPONGSESNENMLYMG 640
Qy 253 RSSEYCLEYFPSSQMLRTGNFTFSYTEFEVPHSSVAHSQSLDRMLNPLIDQYLYLNR 312
Db 641 RTSDAARVNWNHINNNRMNGFNCFGEE--ETKATQNGKLVNTPNGSKDQNRFLTGG 697
Qy 313 QNOSGSAQ-NKDLLFSRGSF-----AGMSVQPKWMLPGPCYRQORVSKTKTDNNNSNF 364
Db 698 TNLNGDLNVEKGTFLSGRTPPHARDIAGISTTKD---PHFTENNEVVVEDDWINRNF 753
Qy 365 TWGASRYNNGRESIIN-----PGTAMASHKDEDEKFFPMSCVMIFGESAG----- 412
Db 754 K--ATTMNVYGNASLYSGRNVANITSNTITASNNAQVHIGYKTDVTCVRSDDYGVYTC 810
Qy 413 ---ASNALD-----NMTIDEDEI--KATNPATERFGTVAVNFQSSSTDPATG 457
Db 811 NSNLSEKALSNFNTNLKGNVNLTEHNSFTLGKANTLGTSITQSVNLKENSHPMLTG 870
Qy 458 DVHMGALPGMWQDRDVLVLOGPIWAKIPHTDGHFHPSPMLGSGGLKNPPOILLKNTPV 517
Db 871 NSNV-----NQLNLNTHGIH-----LNQNDANKVTYVTTLT 902
Qy 518 PANPAPFASATKFSF-----ITQYSTGVASVEIELOKENSKRANPEVQVTSNY 568
Db 903 VNSLSGNGSYFYWDFTNNKSNKVNVNKSATGNFTLOV--ADKTGEPNHNETLTFDASN 959
Qy 569 AKSANDVFTVDN-----NGLYTEPRP-----IGTRYLTRP 598
Db 960 ATRNNLEVTLANGSVDRGAMKYKLRNVNGRDLYNPEVEKNGQVDTTNTITP 1012

RESULT 11
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITAKA
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT

ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 3.3%; Score 106.5; DB 15; Length 2364;
Best Local Similarity 20.5%; Pred. No. 9.3;
Matches 96; Conservative 58; Mismatches 158; Indels 157; Gaps 22;

14 QEPDSSSGIKGTGQPAKRLNPGQTGDSVPDQGLGSPATPAVPTMASGGAP 73
DB QEVQVTTGAG-AGQKTKKSWLYRGLSDRTSKADAS-----ATKVTY----- 889
74 MADNNEGADVGNASGNMHCSTWLGDRVITTTSTRTWALPTNNHLKYQISASSTGASND 133
DB 890 -----VDDGDGNNYTDSDSWLSGRILSTSLRD-----DTGTSHE 922
134 NHFYGYSTPGYDPFNRHGFHSPRDMQRLNNMGRPRKLNPKLENIQVETTTNDGV 193
DB 923 RYHKKY-----WDHNTAYDGLPDAKF-----VREKETTTTKV 956
194 TTIANNLSTVOVPSDEYQLPYVIGSAHQCLPPPADYFMIPOYGYTLNNGSQAVGR 253
DB 957 SSGMRRTVETE-YDTR-----GASTTGPLMRTD-----WGQSSVSD-----NR 997
254 SSFYCLEYFSPQMLRTGNNFTFSYTFBEVPPH-----SSYASQSLDRMLNPLIDQYLYL 309
DB 998 CTTYGRAVNTDNDVSTGAQ-RWTVLQDVVHYSVGCSSIASDSKQ-DGVASTLYD----- 1049
310 NRTQNGSGAQNKDLFSSPAGMSVQPKWMLPGCYRQORVSKYTKTDNNNSFTTGA 369
DB 1050 -----MAISVDANKPV-----DANPT-----ESRITYKSGSYRSTSG- 1082
370 SKYNLNGRESIINGTAMASHKDEDEKF-----PPMSGWIFGESAGA--SNTALDNM 422
DB 1083 --YDAGRVWMSBDG---SHNRLTKYSPANTWPLNGVITTPDPGALPAHAL----- 1132
423 ITDEEIKATNPVATERFTVAVNFQSSSTDPAVDHAGALPGWVQ 471
DB 1133 -----TSTAMTSRFWKGKPTSIDANGVWTMSLDAAGRLV-EVWR 1171

RESULT 12
US-10-062-254-254
; Sequence 254, Application US/10062254
; Publication No. US20020138882A1
; GENERAL INFORMATION:
; APPLICANT: Canoon, Edgar B
; APPLICANT: Canoon, Rebecca E
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Pang, Yiwen
; APPLICANT: Hantke, Sabine S.
; APPLICANT: Lee, Jian-Ming
; APPLICANT: Li, Zhongsen
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Morgante, Michele
; APPLICANT: Niu, Xiping
; APPLICANT: Odell, Joan
; APPLICANT: Sakai, Hajime
; APPLICANT: Zheng, Peizhong
; APPLICANT: Zhu, Qun
; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved in Plant Metabolism
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/062,254
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/630,346
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/146511
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/156006
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/156899
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 60/157287

;; PRIOR FILING DATE: 1999-10-01
;; PRIOR APPLICATION NUMBER: 60/169767
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 60/171054
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: 60/172958
;; PRIOR FILING DATE: 1999-12-21
;; PRIOR APPLICATION NUMBER: 60/171515
;; PRIOR FILING DATE: 1999-12-22
;; PRIOR APPLICATION NUMBER: 60/173535
;; PRIOR FILING DATE: 1999-12-29
;; NUMBER OF SEQ ID NOS: 375
;; SOFTWARE: Microsoft Office 97
;; SEQ ID NO 254
;; LENGTH: 420
;; TYPE: PRT
;; ORGANISM: Zea mays
US-10-062-254-254

Query Match 3.3%; Score 106; DB 15; Length 420;
Best Local Similarity 20.4%; Pred. No. 0.78;
Matches 76; Conservative 54; Mismatches 141; Indels 102; Gaps 19;

44 SVEDPQP-----LGEPPATPAVG---PTMASGGAPMADNNEGADVGNASGNMHC 93
DB 54 AAPPPPPQAKHPAAGPPPPQAGARRPTTPSSSGGNGCEENK----- 99
94 DSTWLDGDRVITTTSTRTWALPTNNHLKYQISASSTGASNDNHFYGYSTPGYDPFNRHFC 153
DB 100 -TIWVD-----LQYMDENYLNCFGPSGEVITKVINRROTQSGEGYVEF----- 147
154 HSPRDMORLNNMGP-----RPKRLNPKLENIQV--EVTNDG--VTIANNLTS- 202
DB 148 -FSHAAEKALQWPTGHWNPNTDPRKLNWASYSMEKSEVASDSIFVGDLAADVTDE 206
203 -TVQVPSDEYQLPYVIGSAHQCLPPPADYFMIPOYGYTLT---NNGSQAVGR-SSFY 257
DB 207 MLELPS-SKYR-----SVKAKVITDANTGRSGYGVVRGDDNDSHAMSENNGVY 258
258 C---LEYPSPQMLRTGNNFTFSYTFBEVPPHSSVASHQ-----SLDRMLN-- 299
DB 259 CSTPRIRIGPATPRRSSGD-----SGSSTPGHSDGDSNRTVYVGLDPPVSEDELKAF 313
300 -----PLDQ--YLYINLRTQNGSGAQNKDLFSSPAGMSVQPKWMLPGPCY 347
DB 314 AKYDLASVKIPLKQKQGFVQFVSRIDAERA-----LQGLNGSLIGKQAVRLSWRSPSH 367
348 RQQRVSKTKTDNN 360
DB 368 KQSRGDSVNRNN 380

RESULT 13
US-10-046-583A-1
; Sequence 1, Application US/10046583A
; Publication No. US20020168743A1
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION OF MODULATORS OF DEOXYXYLUULOSE 5-PHOSPHATE
; FILE REFERENCE: 2037 US Divisioanal
; CURRENT APPLICATION NUMBER: US/10/046,583A
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/626,589
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

LENGTH: 717
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-046-583A-1

Query Match 3.2%; Score 104.5; DB 15; Length 717;
Best Local Similarity 20.4%; Pred. No. 2.4;
Matches 77; Conservative 54; Mismatches 121; Indels 125; Gaps 18;

187 VTNDGVT---IANNLTSTVQVFSDEYQLPYVLSAHQCLPPFPADVPMIQYILT 243
12 IITGGSLTSDCKSTSSSSRLVTD---LP-----SPCLXP----- 45
244 LNNQAVGRSSFCCL-----EYF-----PSOMLRTGNPFSTFEEVPHSSVAHSOS 293
46 -NNNSHS-NRPAKVCASLAKEGEIYSNRPPLDITN-----YPIHMKRLSYKE 93
294 LDRLNPIDQYLYLNTQNSGSA-----QNKDLLFSRGSPPAGMSV 336
94 LKQJSDLRSDVIFNVSKTGHLSGLGVELVVALHYIFNTPODKILM-----DVGHQS 148
337 QPKMLPFCPCYRQORVSKTIDNNSNFTWTGASKYINLNG---RESIINPGTMAASHKD 393
149 YPHKILTG---RRGMPTRQTNGLSGFTKRGESEHDCFGTHSSTISAGLGMAVGRDL 205
394 EDKFFPMGVMIFGKESAGANTALDNVMTDEBEIKATN-----PVAT-----BRFG 441
206 KKKNNVAVVIGDGMATIGQAYEAMNNAGYLDSMTIVILNKKOVSLEPTATLDGSPFVG 265
442 TVAVNFOSSSTDPAT-----GDVHAMGA-----LPQNV-----MODRD 474
266 ALSASLSLQSNPALRELREYAKGWTKQIGBPWQLAKVDEYARGMISGTGSSLFEBELG 325
475 VYLOGPIWAKLPHTDGH 491
326 LYIIGPV-----DGH 335

RESULT 14
US-09-815-242-5845
Sequence 5845, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5845

LENGTH: 1016
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5845

Query Match 3.2%; Score 103.5; DB 10; Length 1016;
Best Local Similarity 21.5%; Pred. No. 4.9;
Matches 122; Conservative 68; Mismatches 196; Indels 181; Gaps 33;

77 NNEGAGD-----VGNASGWHCDSTWLGDRVITTTSTRTALT---YNNHLKQJSSAS 127
189 NTGADGWFLFKGNAB-EYLTNGGILGDKLVNS-GGFKIDTGYIYSSMKTOKAQ 246
128 TGSNDNHYGYSTPGYFDFNRFHCFSPRDQRLNNMGFRPKRLNFKLNIQVKEV 187
247 QG-----YNGY-----GAF-----YKNDSSG-----NSQWGENIDKS 274
188 TTNDGVTIANNLSTVQVFSDEYQ-----LPYVLSAHQCL----- 225
275 KTN--FLNVADNSTNT---SDGKFGQRLNDVILTFYVASTGMRABYAGKTWETSITDL 328
226 -LPPFPADVPMIQYGLTINNSQAVGRSFCLEFSPQMLRT---GNNFTFS----- 276
329 GLSKNOAYNPLITSSORWGLNOGINANG-----NMRTDLKGEFTFDBAPK 375
277 -----YTFEEVPHSSVAHSQSLDRLNPIDQYLYLNTQNSG---SAON--- 321
376 TITELKKVBEIIFPK-----ERKFPDLAPGTEKVR-BGQGEKTIITPILKNPL 426
322 KDLFSRGSF-AGMSVQPKMLP-----GPCYRQORVSKTYTDNNSNFTWTGASKYN 373
427 TGYIISGEBKEKITDPINELTEYGPETIAPGHRDEFPDKLP-----GEKEE 475
374 LNGRESIINP-GTMAASHDDEKFFPMGVMIFGKESAGANTALDNVMTDEBEIK 431
476 VPKPGIKNPETGVAPPPDVSVTKYGPVKGDSIVEKE-----IPEKERRK 523
432 TNPVA--TERFGTVAVNFOSSST-----DPATGDVHAMGALPGWVMD--RDVYLOGPIW 482
524 NPDLAPGTEKVTREGQGEKTIITPILKNPLTGVIIISGSKSEKITDPINELTEGPEF 583
483 AKLPHTDGHFP-----SPLMGFGQLKNPPQILIKNTVPANPPAESATKFSFIT 535
584 ITRGHRD-EFDPKLPFGKEKEVPKPGIKNPETGDVVR-----PPVD-SVTKYGPV-- 632
536 QYSTGVSVIEIWELOKENSKNRPV 562
633 ---KQDSIVEKE-EIPEKERRKFPDL 655

RESULT 15
US-10-198-070-69
Sequence 69, Application US/10198070
Publication No. US20030109437A1
GENERAL INFORMATION:
APPLICANT: AVERBACK, PAUL
APPLICANT: GEMMELL, JACK
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
TITLE OF INVENTION: CELLS
FILE REFERENCE: 59003.000008
CURRENT APPLICATION NUMBER: US/10/198,070
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/306,161
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/306,150
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/331,477
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 69
LENGTH: 1265

TYPE: PRT
ORGANISM: Homo sapiens
US-10-198-070-69

Query Match 3.28; Score 103.5; DB 15; Length 1265;
Best Local Similarity 19.08; Pred. No. 6.7;
Matches 114; Conservative 68; Mismatches 180; Indels 237; Gaps 30;

```
QY 95 STWIG---DRVITSTRTWALPTNNHLYKQI-----SSASTGA 130
Db 49 STFLGILENHLLKKFQVYTWEL--HNKHLFENVFSEPLQSNLPALVQIRLGTTHDTC 106
QY 131 SUNDHYFGYSTPWGYDPNRFHGFSPRDW-----QRLNNMG- 169
Db 107 SEDT---YSTLLQRYRSEBELRVAEMLECOKRIDAVVDEGMTKTKQRMLEDWEL 162
QY 170 FRPRKLNFKLVQKAVTNDGVTYIANNLTSTVQVFSDEYQLP-----YVLGSAHQ 223
Db 163 FKQRRF-----IEQLTNKKAVENTGNNFTDTRHVLSSRLSMPCPNRYRRACAD 214
QY 224 GCL-----PPPADVF-----MIPQGYL-----TLNNGSAVGRSSFYCL 259
Db 215 DCSLSHLITGIMDPPTDIDHQLPLQYDPADYLAERSPPSVSSASSGSSSPTTI 274
QY 260 EYFPPQMLR-TGNNFTFSYTFEEV-PPHSSVAHSQSLDRMLNPLIDQYLYLNRTONQSG 317
Db 275 QQHRRLITDGSAPFTCSDDDEVAPLSAKFAD-----IYPLSNYDDEV 319
QY 318 SAONKDLFRSGSPAGMSVQPKWLPQPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGR 377
Db 320 VA-NMNGIHSELNGGEMALKDESP-----QISST----- 349
QY 378 ESTINPTAMASHKDDCKEFPMSGVMIPEGESAGASNTALDNVMTDE---EELKATNP 434
Db 350 -----SSSSSEADDEADGESSG-----EPPGAPKE--DVLGSRSPRTBESKADSP 394
QY 435 VATERFGTVAVNFQSSSTDPATGVDH-----AMGALP---GMVWQDRDLYL 477
Db 395 PPS--YPT-----QQAQEAQNTCECHVCKQEAASGLTPSANTAGALPRGHQFLSPEKPTH- 446
QY 478 QGPIWAKIPTDGH--FH-----PSPLMGAPGLKNPPQILINKTPVNPANPPA----- 523
Db 447 --PALHLPHIHGHVPLHVPPLPRPLIHPTLYATPP-----FTHSKALPPAPVQNHNTN 498
QY 524 -----EFSATKPAFTITQYSTGVSVIEWELQKNSKRNPP 560
Db 499 KHQVFNASLQDHIYPPSCFGNTPEWNSKSLTSL-----WGSEVMNDKKNMP 543
```

Search completed: July 17, 2003, 18:51:16
Job time: 31.8443 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 21.473 Seconds
(without alignments)
2681.720 Million cell updates/sec

Title: US-09-807-802a-15

Perfect score: 3251
Sequence: 1 TAPGKRPVQSPQEPDSSS.....NNGLYTEPRPIGRVILTRPL 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: PIR 73: *
2: p1r1: *
3: p1r2: *
4: p1r3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	67.0	504	1 VCPV3A	coat protein - ade
2	1701	52.3	732	2 SS2210	coat protein VP1 -
3	556	17.1	673	1 VCPVH5	coat protein VP1 -
4	495	15.2	781	1 VCPV19	coat protein VP1 -
5	258.5	8.0	723	1 VCPVPP	coat protein VP1 -
6	254.5	7.8	729	1 VCPVNA	coat protein VP1 -
7	254.5	7.8	729	1 A60006	coat protein VP1 -
8	244	7.5	722	1 VCPVW2	coat protein VP1 -
9	234	7.2	587	1 B44276	coat protein VP1 -
10	216	6.6	716	1 VCPV2M	coat protein VP1 -
11	216	6.6	727	1 VCPV1F	coat protein VP1 -
12	214.5	6.6	718	1 VCPVIM	coat protein VP1 -
13	214	6.6	722	1 VCPVME	coat protein VP1 -
14	209	6.4	727	1 VCPVFP	coat protein VP1 -
15	208	6.4	737	1 VCPVCD	coat protein VP1 -
16	204	6.3	748	1 VCPVCP	coat protein VP1 -
17	198	6.1	584	2 S49594	capsid protein VP2
18	192	5.9	722	1 VCPVON	coat protein VP1 -
19	191	5.9	702	1 VCPVAP	coat protein VP1 -
20	143.5	4.4	648	2 S50856	wnm protein - rat
21	129	4.0	1072	2 A86827	hypothetical prote
22	123.5	3.8	690	2 AB0124	probable Tomb-depe
23	123	3.8	1142	2 T37455	enamelin precursor
24	122.5	3.8	880	1 SYBSYS	valine-tRNA ligase
25	121	3.7	635	2 P96660	protein F2K1.10 (
26	121	3.7	1113	2 S28925	nuclear pore compl
27	120.5	3.7	931	2 T49710	related to glucan
28	119	3.7	667	2 A41311	transcription fact
29	118.5	3.6	1180	2 E86719	hypothetical prote

30	118	3.6	739	2 T52289	probable transketo
31	116.5	3.6	642	1 S34416	transcription fact
32	116.5	3.6	1338	2 T30565	MAP kinase kinase
33	115.5	3.6	345	1 G97024	probable phospho
34	114	3.5	956	2 T08144	myrosinase-binding
35	114	3.5	2271	2 P90073	hypothetical prote
36	113.5	3.5	2817	2 B97033	uncharacterized pr
37	113	3.5	1071	2 B85343	hypothetical prote
38	112.5	3.5	666	2 T52548	class A helix-loop
39	112.5	3.5	3078	2 T28432	variant-specific s
40	111	3.4	1777	2 T34369	hypothetical prote
41	110.5	3.4	655	1 ALKBG	cytoplasmic dectri
42	110.5	3.4	1742	2 T17120	cellulase (EC 3.2.
43	110	3.4	868	2 A82515	conserved hypochet
44	110	3.4	1296	2 C82521	hemolysin-type cal
45	110	3.4	2825	2 T14271	Doc4 protein, stre

ALIGNMENTS

RESULT 1

VCPV3A
coat protein - adeno-associated virus type 2

C:Species: adeno-associated virus type 2
C:Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jul-1999

C:Accession: A03698
R:Stivastava, A.; Lusby, E.W.; Berns, K.I.
J. Virol. 45, 555-564, 1983

A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.
A:Reference number: A03694; PMID:83164299; PMID:6300419

A:Accession: A03698
A:Status: translation not shown
A:Molecule type: DNA

A:Key words: coat protein

Query Match 67.0%; Score 2177.5; DB 1; Length 504;
Best Local Similarity 82.5%; Pred. No. 7.9e-142;
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

QY	66	MASGGGAPMADNNEGADGVGNASGNHCDSTWIGDRVITTSRTFMALPTYNHLYQIIS	125
DB	1	MATGGAPMADNNEGADGVGNSSGNHCDSTWIGDRVITTSRTFMALPTYNHLYQIIS	60
QY	126	ASTGASNDNHYFGYSTPWGYFDFNRFHCHSPRDWQRLINNMGFPKRLNFKLFINIQV	185
DB	61	QS-GASNDNHYFGYSTPWGYFDFNRFHCHSPRDWQRLINNMGFPKRLNFKLFINIQV	119
QY	186	EVTNDGVTTIANNLSTVOVPSDSEYOLPYVIGSAHQGLPPPADVEMIPQYGYITLN	245
DB	120	EVTNDGVTTIANNLSTVOVPSDSEYOLPYVIGSAHQGLPPPADVEMIPQYGYITLN	179
QY	246	NGSOAVGRSSFFCYLXEPSPQMLRTGNFFSYFEFVPHFSYAHOSQDLRLNPLIDQY	305
DB	180	NGSOAVGRSSFFCYLXEPSPQMLRTGNFFSYFEFVPHFSYAHOSQDLRLNPLIDQY	239
QY	306	LYVINRTONQSGSAOKDILFSGRSPAGMSVQPKNMLPGFCYRQORVSKTKTNNNSFT	365
DB	240	LYVLSRTNTPSGTTSQRLQFSGAGASDIRDQSRNMLPGFCYRQORVSKTKSADNNSEYS	299
QY	366	WTGASKYNLNGRSGIINPGTAMASHODEDKFPFMGCVMTFGESKAGASTALDNVITD	425
DB	300	WTGATKYHLNGRSLVNP--AMASHDDEKFFPOSGLVIFGQSGSEKTVNIEKWTID	357
QY	426	EEETIKATNPATATRGVAVNPFQSSSTDPATGVHMGALPGWVODRDVYLGPIWAKI	485
DB	358	EEETIGTNPATATRGVAVNPFQSSSTDPATGVHMGALPGWVODRDVYLGPIWAKI	417
QY	486	PHTDGPHSPPLMGFGGLKNPQIILIKTVPANPAEFSATKFAFIQVSTG	540

Db 418 PHTDGHFHPBPLMGFGFLKHPPOILLKNTVPANPSTTSPASAFSPITQYSTG 472

RESULT 2

552210
coat protein VP1 - muscovy duck parvovirus
N:Alternate names: VP1 protein
C:Species: muscovy duck parvovirus
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-2000
C/Accession: S52210
R:Zadort, Z.; Erdel, J.; Nagy, J.; Kisari, J.
submitted to the EMBL Data Library, September 1993
A:Reference number: S52209
A:Accession: S52210
A:Molecule type: DNA
A:Residues: 1-732 <ZAD>
A:Cross-references: EMBL:X75093; NID:G609091; PID:CAAS2984.1; PID:G609093
A:Experimental source: strain FM
C:Genetics:
A:Gene: VP1
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein

Query Match 52.3%; Score 1701; DB 2; Length 732;
Best Local Similarity 53.0%; Pred. No. 7.1e-109;
Matches 325; Conservative 83; Mismatches 165; Indels 40; Gaps 11;

QY 6 KRPEQSPQEPDSSSGIGTK-TGQOPAKKRLNFGQTGDSSESVPDPPLG-----E 53
Db 141 EEPVNTAPAKKS-----GKLTDPHYKPKLSE-----ENSPSPSNGSASAAATGSE 192
QY 54 PRPAPAVPTTMAAGGAPWADNNEGADGVGNASGNMHCDSITWLGDRVITSTRVALP 113
Db 193 PVAP-----NMAEGSGAMGDSAGADGVGNASGNMHCDSITWLGDRVITSTRVALP 246
QY 114 TYNHLYKOISSASTGASNDNHYFGYSTPMGYFDFNRFCHFSPRDQRLINNMWGRPRK 173
Db 247 SYNHNWQALITSGNPNPSN-TQVAGYSTPMGYFDFNRFCHFSPRDQRLINNMWGRPRK 305
QY 174 RLNKLNLNIOYKETTNDGVTIANNLSTVQVSDSEYQLPYLGSAGHQCCLPPPADV 233
Db 306 ALKKEIRNVOYKEVTTDDQTKTIANNLSTIQITFDEHQLFYVLGSATBCTMPPEPSDV 365
QY 234 FMIPQYGLTLN---NGSOAVGRSSFCLEFYFSPQMLRTGNFTFSTFEVPPHSSVAH 290
Db 366 YALPQYCTMHTNOSGARFDRSAFYCLEFYFSPQMLRTGNFTFSTFEVPPHSSVAH 425
QY 291 SOSIDRLMNPILDOYLYLNTNOSGSAQNKDLLFSRGSPPAGMSVQPKMLPPCYRQ 350
Db 426 SOSIDRLMNPILDOYLYLNTNOSGSAQNKDLLFSRGSPPAGMSVQPKMLPPCYRQ 479
QY 351 RVSKTK---TNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKPPMSVMIK 408
Db 480 RVKRVSGSTDYVNAWSIMSKNKFLLKXREYLLDPGVATHTTEDQASSVPAQNIISIAK 539
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTPATGDVYAMALP 466
Db 540 DPKRSSGTLAGISIMTDEBEIAPNTNGVGRPGYLVNTEQNTTAPTALEVLGALP 599
QY 467 GNVQODRDVYLOGBIMAKIPHTDGHFHPSPILMGSGFGKLNPPQILLKNTVPANPAPFS 526
Db 600 GNVQONDYLYLOGBIMAKIPHTDGHFHPSPILMGSGFGKLNPPQILLKNTVPANPAPFS 659
QY 527 ATKASITQVSTQGVSEIEMELQKNSKRNPEVOYTSNYSASAVDFTVNNGLYTE 586
Db 660 NQKNSYITQVSTQGVSEIEMELQKNSKRNPEVOYTSNYSASAVDFTVNNGLYTE 719

QY 587 PRPIGTATRPPL 599
Db 720 DRLIGTRVLTQNL 732

RESULT 3

VCPV5
coat protein VP1 - bovine parvovirus
N:Contains: coat protein VP2
C:Species: bovine parvovirus
C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C/Accession: A26104
R:Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.
J. Virol. 60, 1085-1097, 1986
A:Title: Complete nucleotide sequence and genome organization of bovine parvovirus.
A:Reference number: A26104; MIMD:87061184; PMID:3783814
A:Accession: A26104
A:Molecule type: DNA
A:Residues: 1-673 <CHB>
A:Cross-references: EMBL:M14363; NID:G333454; PID:AA859847.1; PID:G808805
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
C:Keywords: product: coat protein VP2 #status predicted <VP2>

Query Match 17.1%; Score 556; DB 1; Length 673;
Best Local Similarity 25.0%; Pred. No. 2.6e-30;
Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;

QY 1 TAPGKRPRVQSPQEPDSSSGIGTKGQOPAKKRLNFGQTGDSSESVPDPPLGSP--PAP 58
Db 100 TSKGGRALRKLYFARSNKGAKKARREPAPSTSNQONNEVSDINDEAGNPILATR 159
QY 59 AAVGPTTMAAGGAPWADNNEGADGVGNASGNMHCDSITWLGDRVITSTRVALP 118
Db 160 SVVSGSGVGGG-----RGSSGVGYSTGWTGTTSENVTKNTRQFLCIDIKNGH 211
QY 119 LVYKOISSASTGASNDNHYFGYSTPMGYFDFNRFCHFSPRDQRLINNMWGRPRKLNFK 178
Db 212 LVYS-BVLTNGDFAHQY-AITPMGYFNNOYSSHFSPRDQRLINNMWGRPRKLNFK 269
QY 179 LENIOYKETTNDGVTIANNLSTVQVSDSEYQLPYLGSAGHQCCLPPPADVFMIP 237
Db 270 VVNLQIKQITDAMGVYVNNDLTAGMHIKCDSDHYPPYVQHPWDQCEMELPNSIWEIP 329
QY 238 QYGYLT-----NGSOAVGRS---SFYCLEFYFSPQMLRTGNFTFSTFEVPPHS 286
Db 330 QYVITPAPISVDNNTNTNVEEHLKGVPLMYLENSDHEVLRNG----- 373
QY 287 SVASOSLRLMNPILDOYLYLNTNOSGSAQNKDLLFSRGSPPAGMSVQPKMLPPCYRQ 331
Db 374 -----RIYIYIQLWRLBMDRKHNIQHASDDVQSTGQKQKULLQRTQPKQRF 424
QY 332 AGMSVQPKMLPPCYRQQRVRSKTKTDNNSNFTWTGASKYNLNGRESIINPGTAMASHK 391
Db 425 QNAAKRTSNMMSGP-----GIARGTINATLQTSAGALVTMT 462
QY 392 DDEDKFFPMGVM-----IFGKE-----SAGASNTALDNVMTDE 426
Db 463 NGAD---VSGVAAVGVSTDPITYGQOPESLRLYRSASABEQNPILN----- 512
QY 427 EEIKATNPVATERFGTVAVNFQSSSTPATGDVYAMALP 466
Db 513 -----AARHFTREARLKITGSGADGVKEMWMLPNQWMDSPISRYPYIWKVP 564
QY 487 HTDGHFHPBPLMGFGFLKHPPOILLKNTVPANPAPFS 526
Db 565 RVNRKTLTLDQDSISIMSHPPGTFIKLARIPVPGND-----SFLNIYVTVQVSC 615
QY 545 EIEMLQKNSKRNPEVOYTSNYSASAVDFTVNNGLYTE 584
Db 616 EVVWEVEKGTGKMRPEYMHSS---ATNMSVDATYINNAGY 653

RESULT 4
VCPV19
coat protein VP1 - parvovirus B19 (strain Au)
C:Species: parvovirus B19
A:Note: host Homo sapiens (man)
C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999

RESULT 3

C/Accession: A24299
 R/Shade: R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.
 J. Virol. 58, 921-936, 1986
 A/Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated from
 A/Reference number: A24299; MUID:86200451; PMID:3701931
 A/Accession: A24299
 A/Molecule type: DNA
 A/Residues: 1-781 <SHA>
 A/Cross-references: EMBL:U13178; NID:G333375; PID:AAA6667.1; PID:G333377
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein

Query Match 15.2%; Score 495; DB 1; Length 781;
 Best Local Similarity 27.3%; Pred. No. 5e-26;
 Matches 148; Conservative 84; Mismatches 257; Indels 54; Gaps 16;

QY 51 LGEPPTAAVPTTMAAGGAPMADNNEGADGVGNAGNHCDSWTLDGRVITSTRTW 110
 DB 213 LPEVPAYNAEKYPSMTSVNSAB-ASTAGAGGGSNSVKSMMBSGATPSANSVCTESRCP 271
 QY 111 ALPTNNHLTKOISSASTGASND-----NHVYGSTPWGYPEFNFPHCHSPRDM 160
 DB 272 LIPYDEPHHYKVSPPASCHNAGSKAKVCTTSPWGYSTPWKYLDFNAINLFPSPLEF 331
 QY 161 QRLINNMWGFPRKLNFKLNIOYKEVT--TNDGVTTIANNLSTVOVFSDEYQLPYVL 218
 DB 332 QHLEIVNGSLAPDALVTITSEIAVKDVTDKTGGV-QVTDSTTGRLCMLVDHEKYRYVL 390
 QY 219 GSAHQGLPPPADVEMIPQYGYLTLLN-GSQAVG-----RSSFCLEYFPSSQMLR 268
 DB 391 GQGDITLAPLPIWVYPPQYAVYLVGDVNTQGISGDSKLLASBESAFVLEHSSFQLLG 450
 QY 269 TGNNTFFSYFEFEPFSSSYAHSQSLDRMLNPLIDQVLYLNTQNGSSAOKKDLFSR 328
 DB 451 TGTASTSYFPPVPPEPNELECCSHFEYEMNPL--TGSRLGVPTLGGPKPRSL-- 503
 QY 329 GSPAGNSVQPKMLPGPCYRQQRVSKTKTNNNSNFTWGTASKYNLGRSEIINPG-TAM 387
 DB 504 -THEDHALQPNMPGFLVNSVSTKEGDSSTGAKRLTGLSTGTSQNTSLRPGVSG 562
 QY 388 ASHKDEDEKFPMSGVMI FKESSAGASNTALDNV-----MITDEEIKATNPATERF 440
 DB 563 PYHMDTKVYTGAINASHQDTYG---NAEDKEYQGVGRFPNEKQLKOLGLMHTY 619
 QY 441 GTAAVNVQSSSTPATGDVAMGALPGMWQDRVYVIGRPIMAKIPIHTDHPHS-PLMG 439
 DB 620 -----FNNKGTQGYTDQIE-RPLMVGSVNNRRALHESQLMSKIPLADSFRTQFALG 672
 QY 500 GFGLNKPPQIILKMTVPANPAPAEFATKFASTTQYSTGOVSVEIEMEL-QKENSRRW 558
 DB 673 GWMGHQPPQIFLK--ILPQSGPIGIGIKSMGITTLVQYAVGIMVTMTFTLGRKATGRV 730
 QY 559 NPE 561
 DB 731 NPQ 733

RESULT 5

VCEVP
 coat protein VP1 - porcine parvovirus
 N/Contains: coat protein VP2
 C/Species: porcine parvovirus
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 12-Apr-1996
 C/Accession: B33302
 R/Ranz, A.I.; Mancius, J.J.; Diaz-Aroca, E.; Casan, J.I.
 J. Gen. Virol. 70, 2541-2553, 1989
 A/Title: Porcine parvovirus: DNA sequence and genome organization.
 A/Reference number: A33302; MUID:90010964; PMID:2794971
 A/Accession: B33302
 A/Molecule type: DNA
 A/Residues: 1-723 <RAN>
 A/Cross-references: EMBL:D00623
 C/Genetics:

A/Intons: 10/1
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein
 F/45-723/Product: coat protein VP2 #status predicted <VP2>

Query Match 8.0%; Score 258.5; DB 1; Length 723;
 Best Local Similarity 23.1%; Pred. No. 7.7e-10;
 Matches 155; Conservative 92; Mismatches 270; Indels 153; Gaps 35;

QY 9 VEQSPQEPDSSSGITG-----TGQOPAKRLNFCQTGDSVPPQIPGEPAIP 58
 DB 103 VRSPPKHGSKPPGRPARRHIFINLAKKAGTSNTNSNSSEVDEHPNIP--AAE 160
 QY 59 -AAVGTMAAGGAPMADNNEGADGVGNAGNHCDSWTLDGRVITSTRTW 112
 DB 161 LSAITNESGGGGGG-----GKAGAGVGVSTGSPNNQTFQYIGEGVATITAHASLIHL 216
 QY 113 PTYNNHLTKOI-----SSASTGAS-NDNHYFGYSTPWGYPEFNFPHCHSPRDMRLINN 166
 DB 217 NMPEHTYKRIHVLNSESAGAGMVDADHTQVTPWMLIDANAGVWFPNADWQLISNN 276
 QY 167 NMGFRKLNFKLNIOYKEVT--TNDGVTTIANNLSTVOVFSDEYQLPYVLSAHO 223
 DB 277 MTEINLVSEBOALFNVLKTTITESATSPPTKYNNDLTASLWALDPTNNTLPYTPAAPS 336
 QY 224 GCLPPPADVEMIPQYGYLTLLN-GSQAVG-----TANNGSQAVERSS-----FYCLE-YFP 263
 DB 337 ETIGFYPLPTKPTQRYTLYLSCIRNLNPTLYGQSPNNRLNRLHSDIMEFTTENAVP 396
 QY 264 SQMLRTGNPF-TFSYFEFEPFSSSYAHSQSLDR-----LMNPLI--DOVLYIANRT 312
 DB 397 IHLRTGDESGTGIYHFDTRPL--KLTHSQNNRSGLPPLKLTETTEGDDHPTLPAA 454
 QY 313 QNOSGAOKKDLFSRSGSPAGNSVQPKMLPGPCYRQQRVSKTKTNNNSNFTWGTASKY 372
 DB 455 NTRKGHYQTINNSYATAT--AIRP-----AQGVNTPYNNPEYSGNPGF 496
 QY 373 NNGRESIINP--GTMAHKDEDEKFPMSGVMI FKESSAGASNTALDNV-----NMVITDEE 427
 DB 497 -----LPIVPIADTQYNDDEPN-----GAIRTMQYQHGHLTSSQ 533
 QY 428 EIK-ATNP-----VATERFGTV-VNFOSS-----TDPATG--DVHAMGALP---- 466
 DB 534 ELERYTFNPOSKGRAPKQPFQDAPLNTNNGTLPLSDPIGKSNHFNMTLWTYGP 593
 QY 467 -GMWQDRVYVIGRPIMAKIPIHTD--GHPHSPLMGGFLK-NPPQIILKMTVPANP 522
 DB 594 LTLNNTAVFPVPGQIMDELDITDKPRLH--VTAPFVCKNNPPQOLFVKIAP--NLT 647
 QY 523 AEPSEA-TKFASTTQYSTGOVSVEIEMELQKENSRRWNEVOYTSYAKSANDFTVDNN 581
 DB 648 DDFNADSPQOPRIITYNSNFMWKGTLLTKAKRSSNNMNPIDQHTT-----TAENI 697
 QY 582 GLYTEPRPIG 591
 DB 698 GKVI-PTNIG 706

RESULT 6

VCEPVNA
 coat protein VP1 - porcine parvovirus (strain NADL-2)
 N/Contains: coat protein VP2
 C/Species: porcine parvovirus
 C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
 C/Accession: B33743; D48472
 R/Vasudevarchara, J.; Basak, S.; Srinivas, R.V.; Compans, R.W.
 Virology 173, 368-377, 1989
 A/Title: Nucleotide sequence analysis of the capsid genes and the right-hand terminal pa
 A/Reference number: A33743; MUID:90085785; PMID:2596019
 A/Accession: B33743
 A/Molecule type: DNA
 A/Residues: 1-729 <VAS>
 A/Cross-references: GB:M32787; NID:G332983; PIDN:AAA46917.1; PID:G332985

R:Bergeron, J.; Menezes, J.; Tijssen, P.
Virology 197, 86-98, 1993
A:Title: Genomic organization and mapping of transcription and translation products of
A:Reference number: A48472; MUID:94025614; PMID:8212598
A:Accession: D48472
A:Molecule type: DNA
A:Residues: 11-729

A:Experimental source: MADL-2, ATCC VR-742
A:Note: sequence extracted from NCBI backbone (NCBIN:138789, NCBIPI:138794)
C:Genetics:
A:Insertions: 10/1
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein; glycoprotein
F:151-729/Product: coat protein VP2 #status predicted <VP2>
F:32,172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 7.8%; Score 254.5; DB 1; Length 729;
Best Local Similarity 22.7%; Pred. No. 1.5e-09;
Matches 152; Conservative 94; Mismatches 272; Indels 151; Gaps 34;

```
QY 9 VEOPOEDSSSGIGK-----TGQOPAKKRLNFGQDSESVDPDPLGEPATP 58
D 109 VRSPRKHPSKPPGKRPAPRHIFINLAKKAKGTSTNSMSSENEQHPINAGTEL- 167
QY 59 AAVGPTTMAAGGAPMADNNEGADGVGNASG--NMHCDSTWIGD--RVITTSRTWALP 113
D 168 SATNGESGGGGGGG---GGRGAGGVGVSTGFNNQTEFOYLGEGVIRITAHASRLHLN 223
QY 114 TYNHLYKOIS--SASTGAS---NDNHFGYSTPMGYFDNRPHCHFSRPMQRLINN 167
D 224 MPEHETRYKRIHLVNSSSGVAGQVODDAHTQVTPWLSLDANAGVFNFPADQOLISNM 283
QY 168 WGFPRKLNFLFNIOVKEVT---TNDGVTTIANLNTSTQVPSDSEQLPYVLSAHOQ 224
D 284 TEINLVSEFQETFNVLKTTTESATSPPTKIYNDLTLASLWALDINNLTLYTPAARPE 343
QY 225 CLPPPADVFMIPQYGYL-----TLNNGSQAVGNS-----SFYCLE-YFPS 264
D 344 TLGFYFPLPTKPYQRYVYLSTCNLNPPTVYTGSGQOITDSIQGLHSDIMFYTIENAVP1 403
QY 265 QMLRTGNF--TFSYTFEEVPHSSVHSSQSLDR-----LNPPLI--DOYLYLNRQ 313
D 404 HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRSIGLPRKLTLEPTTEGQHPGTLPAAN 461
QY 314 NQSSAQNKDILFRSGSPAGMSVQPKMWLPFCYRQORVSKTIDNNNSFTWTGASKYN 373
D 462 TRGYHOTINNSYTEAT---AIRP-----AQVGYNTPYNNFEYSNGGPF- 502
QY 374 LNGRESIINP--GTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD--NVMITDEE 428
D 503 -----LTPIVPTADTQYNDDEPN-----GAIKFTMDYQHGHLTSSQE 540
QY 429 IK--ATNP-----VATERFGTVA--VNFQSS-----TDPATG--DYHAMGALP----- 466
D 541 LERTYFNPQSKGAPKQOFPQOAPLNLNTNNGTLPSDPIGKSNMHNFNNTLNTYGP 600
QY 467 GWMVQODRVYLGPIWAKIPIPTD--GHFHPSPLMGGFGLK--NPPQOLIKNTPVANPAEFA 523
D 601 TALNNTAPVFNPGQIMDELDLTKPRH---VTAFFCKANNPQGLVXKAP---NLTD 654
QY 524 EFSA-TKFASTIYOSTGQVSEIEMELQKENSKRANDEVOTYSNYAKSANVDFVDNNG 582
D 655 DFNDSPOQPRILITYSNFMWKGTLTFTAKMRSSNMWNP1QOHTT-----TAENIG 704
QY 583 LYTEPRPIG 591
D 705 NYI-PTNIG 712
```

RESULT 7
A60006
coat protein VP1 - porcine parvovirus (strain 90HS)
N:Contains: coat protein VP2

C:Species: porcine parvovirus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C:Accession: A60006
R:Skutvold, W.; Nishimori, T.; Ushimi, C.; Nakajima, H.
Virus Res. 13, 79-86, 1989
A:Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.
A:Reference number: A60006; MUID:89319168; PMID:2750278
A:Accession: A60006
A:Molecule type: DNA
A:Residues: 1-729 <SAK>
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein; glycoprotein
F:151-729/Product: coat protein VP2 #status predicted <VP2>
F:172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 7.8%; Score 254.5; DB 1; Length 729;
Best Local Similarity 22.7%; Pred. No. 1.5e-09;
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

```
QY 9 VEOPOEDSSSGIGK-----TGQOPAKKRLNFGQDSESVDPDPLGEPATP 58
D 109 VRSPRKHPSKPPGKRPAPRHIFINLAKKAKGTSTNSMSSENEQHPINAGTEL- 167
QY 59 AAVGPTTMAAGGAPMADNNEGADGVGNASG--NMHCDSTWIGD--RVITTSRTWALP 113
D 168 SATNGESGGGGGGG---GGRGAGGVGVSTGFNNQTEFOYLGEGVIRITAHASRLHLN 223
QY 114 TYNHLYKOIS--SASTGAS---NDNHFGYSTPMGYFDNRPHCHFSRPMQRLINN 167
D 224 MPEHETRYKRIHLVNSSSGVAGQVODDAHTQVTPWLSLDANAGVFNFPADQOLISNM 283
QY 168 WGFPRKLNFLFNIOVKEVT---TNDGVTTIANLNTSTQVPSDSEQLPYVLSAHOQ 224
D 284 TEINLVSEFQETFNVLKTTTESATSPPTKIYNDLTLASLWALDINNLTLYTPAARPE 343
QY 225 CLPPPADVFMIPQYGYL-----TLNNGSQAVGNS-----SFYCLE-YFPS 264
D 344 TLGFYFPLPTKPYQRYVYLSTCNLNPPTVYTGSGQOITDSIQGLHSDIMFYTIENAVP1 403
QY 265 QMLRTGNF--TFSYTFEEVPHSSVHSSQSLDR-----LNPPLI--DOYLYLNRQ 313
D 404 HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRSIGLPRKLTLEPTTEGQHPGTLPAAN 461
QY 314 NQSSAQNKDILFRSGSPAGMSVQPKMWLPFCYRQORVSKTIDNNNSFTWTGASKYN 373
D 462 TRGYHOTINNSYTEAT---AIRP-----AQVGYNTPYNNFEYSNGGPF- 502
QY 374 LNGRESIINP--GTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD--NVMITDEE 428
D 503 -----LTPIVPTADTQYNDDEPN-----GAIKFTMDYQHGHLTSSQE 540
QY 432 TNP-----VATERFGTVA--VNFQSS-----TDPATG--DYHAMGALP----- 471
D 546 FNPQSKGAPKQOFPQOAPLNLNTNNGTLPSDPIGKSNMHNFNNTLNTYGP 605
QY 472 DRDVLQGPIMAKIPIPTD--GHFHPSPLMGGFGLK--NPPQOLIKNTPVANPAEFA- 527
D 606 TAVVFNPGQIMDELDLTKPRH---VTAFFCKANNPQGLVXKAP---NLTD 659
QY 528 TKFASTIYOSTGQVSEIEMELQKENSKRANDEVOTYSNYAKSANVDFVDNNG 587
D 660 SPOQPRILITYSNFMWKGTLTFTAKMRSSNMWNP1QOHTT-----TAENIG 708
QY 588 RPIG 591
D 709 TNIG 712
```

RESULT 8
VCPV2
coat protein VP1 - parvovirus H1
C:Species: parvovirus H1
A:Note: host Homo sapiens (man)

C/Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
 C/Accession: A03699
 R/Rhodo III, S.L.; Paradiso, P.R.
 J. Virol. 45, 173-184, 1983
 A/Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid
 A/Reference number: A03695; MUID:83112183; PMID:6823009
 A/Accession: A03699
 A/Molecule type: DNA
 A/Residues: 1-722 <RHO>
 A/Cross-references: EMBL:X01457; EMBL:J02198
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein

Query Match 7.5%; Score 244; DB 1; Length 722;
 Best Local Similarity 21.4%; Pred. No. 7,6e-09;
 Matches 142; Conservative 94; Mismatches 267; Indels 160; Gaps 29;

```

QY 8 PVEQSPQEPDSSGIGTGOQ---PA-----KKRLNFGQYDSESVPPQPLGEP 54
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 82 PKLSTDSBP-GTSGVSRPGRKTPAHIPVNOARAKKRASLAQQRLLTMSDOTETNP 140
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 55 PATPAVGPPTMASGGAPMADNNEGADVGNASGNMHCSTW--LGDVITTTSTRTWAL 112
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 141 DTGIANARVERSDAGGGS-SGGGSGGGGIGVSTGYDNQTYKFLDGDWVEITAHASRL 199
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 113 -----PTVNNHLVQIISASTGASNDNHYFGSTPMGVFDFNRPHCHSPR 158
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 200 LHLGMPSEYECRVYANNQTHGHTKVKGMAYDTHQIWI-TPWSLVVANANGWVFOFS 258
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 159 DWORLNNMGPFRPKRLNFKLNIQVKEVT-----NDGVTIANMLSTVQVFSSEYO 213
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 259 DWQFIQMSMESLNLDSIQELFNVVVKTVEQAGADAIKVVNNDLTACMVALDSNNI 318
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 214 LPLVLSAHGCL-----PPPAD--VPMIROYGLTLNNGSQ-----ANG-----R 253
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 319 LPTPAQTSITLGFYWKPTAPAPRYRYFPMPOULSVTSSNSAEGTQITDTTIGEPOALN 378
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 254 SSFYCLE-YPPSQMLRTGNNFTF-SYFEEVPEFHSYA-----HSQSLDRILM-PLIDQY 305
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 379 SQFTTENTLPIILRLGDSFTTGTITFNTDPLKLTHTWOTNRHLAQITLPLISDTA 438
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 306 LVIYINRTONGSAGNNDLFF-----SRGSPAGMSVOPKMWL 342
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 439 TAILUNGRFSGTQTONVVYVEALTRPAQIGFMOPHNFENRGRPKVVPV-----494
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 343 PEGCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGR--ESTIINPETA-----MASH 350
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 495 -----LDITAGEHDHAN--GAIRFNYSKQHGEDWAKOGAAPERYTWDALDSAG 541
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 391 KODEDKPFPMGSMWIFGKESAGASNTALDVMITDEEIKK-TNPVATERFGTVAVNFOS 449
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 542 KDTARCTV-----QSAPIISIPNONOILQREDAIAGRTMHTVNSIGPLSAF 591
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 450 SSTDPATGVHANGALPGMWODRDVYLQGPIMAK--IPHTDGHFSPPLMGSGFGILKNP 506
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 592 PHEDP-----IYPNGQIMDKELDLHKPRILHTAFAFV---CKNNP 628
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 507 PPGIILKNTPVAPNPAEF--SATKPAFTIYSTGVSVEIEMELQKNSKRNPEVOY 564
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 629 PGGLFHLHGP---NLTDQFDPNSTVSRIVT-YSTPYWKILKFKAKILRNLTMNPFYQA 684
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 565 TSN 567
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 TTD 687
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 9
 B44276
 coat protein VP1 - parvovirus Lu111
 C/Species: parvovirus Lu111
 C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 26-Feb-1999
 C/Accession: B44276
 R/Diffoot, N.; Chen, K.C.; Bates, R.C.; Lederma, M.

Virology 192, 339-345, 1993
 A/Title: The complete nucleotide sequence of parvovirus Lu111 and localization of a unique
 A/Reference number: A44276; MUID:93297126; PMID:8517025
 A/Accession: B44276
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-587 <DIF>
 A/Cross-references: GB:M81888
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein; glycoprotein
 F:49,90,220,304,371,503,511,514,539,571/Binding site: carbohydrate (asn) (covalent) #stat

Query Match 7.2%; Score 234; DB 1; Length 587;
 Best Local Similarity 23.1%; Pred. No. 2.7e-08;
 Matches 137; Conservative 74; Mismatches 231; Indels 152; Gaps 29;

```

QY 59 AAVGPTTMASSGGAPMADNNEGADVGNASGNMHCDS--TWIGD--RVITTTSTRTWAL 112
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 23 AADGPG--GSGGG-----GSGGGGAGVSTGSDNQTHYKFLDGDWVEITAYSTRWVHLN 74
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 113 -PTVNN-----HLVQIISASTGASNDNHYFGSTPMGVFDFNRPHCHSPRDMQRLN 165
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 75 MPKSENYECRVYANNQTHGHTKVKGMAYDTHQIWI-TPWSLVVANANGWVFOFSQVYISN 133
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 166 NMWGPFRPKRLNFKLNIQVKEVT-TNDGVTTIA--NNTSTVQVFSSEYOPLVYLSA 221
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 134 NMHILHSLDQELFNVVVKTVEQNTGAEAIKVVNNDLTAAAMVALDSNNILPTPAID 193
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 222 HQGCL-----PPPPA-----DVPMIROYGLTLNNGSQAVGRSSFYCLEYPP 263
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 194 NQETLGFYWKPTIPSPRYRYFSCDRLNLSVYKDEAGTITDTWGLASGNSQFTIENTQ 253
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 264 S-CMLTGNMF-TFSFEEVPEFHSYSASQSLDRILMNLIDQ-----YL 306
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 254 RINLRTGDEYATGIVTPELRILHTWQNRHLQCPQITELPSSDTANATLTARGYR 313
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 307 YVILNRTON-----QSGSAQNRD-LIFSRSAGMSVOPKMWLPG-----344
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 314 SGLTQIQGNDVTEATRVPAQVFGPHDNFETSAGFPKVPVYADDTIGLGDHANGS 373
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 345 --PCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDDKPFPMG 402
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 374 LRYTYDKHQGQSWASQNNKDRYTW--DAVNYD-SGR-----406
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 403 VMIFGKESAGASNTALDVMITDEEIKKATNPATR--FGTVAVNFOS--SSTDPATG 457
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 407 -----WTNNCFIQSVPTSEPN--ANOILITRMDIAGKTIDHFTNAFNSYGLPLA 454
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 458 DVYAMGALPGMWODRDVYLQGPIMAK--IPHTDGHFSPPLMGSGFGILKNP 514
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 455 FPH-----PAPITPQGIWMDKELDLHKPRILHTQAPFV---CKNNABGOLLVRL 500
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 515 TP--VPANPAEBSATKPAFTIYSTGVSVEIEMELQKNSKRNPEVOYTS 566
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 501 APVLTLQYDVPNSNLSRIIVTYGTFFWKGLTLTKAK--MRPNA-TNNPVFQISA 550
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 10
 VCPV2M
 coat protein VP1 - minute virus of mice
 C/Species: minute virus of mice, murine parvovirus
 C/Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
 C/Accession: A03700
 R/Atcell, C.R.; Thomson, M.; Merchilinsky, M.; Ward, D.C.
 Nucleic Acids Res. 11, 999-1018, 1983
 A/Title: The complete DNA sequence of minute virus of mice, an autonomous parvovirus.
 A/Reference number: A03696; MUID:83143341; PMID:6298737
 A/Accession: A03700
 A/Molecule type: DNA
 A/Residues: 1-716 <AST>
 A/Cross-references: EMBL:V01115
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein

Query Match 6.6%; Score 216; DB 1; Length 716;
 Best Local Similarity 20.8%; Pred. No. 6.3e-07;
 Matches 148; Conservative 93; Mismatches 271; Indels 198; Gaps 34;

```

QY 8 PVEQSPQEPDSSSGIGTKG-----QQPAKKL-----NFQGT-GDSESVDP 48
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 82 PKLATSESP-CTSGVSRAGKRTTRPPAYFIINQARAKKLTSSAAQSSQTSQDTSQPD 140
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 49 QPLGEPPA--TPAAVPTTASGGAPADNNEGADVGNASGWHCHDS--TWIGD---R 101
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 141 GNAAVSAARVERAADPG--GSGGQ-----GSGGGGVGVSTGSDYDQTHRRFLGDGWE 192
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 102 VITSTRWALPTNNHLYKQI-----SSASTGASNDNHYFGYSTPMGFYDNRPHCH 154
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 193 ITALATLVLHLMKSENYCRIRVHNTDTSVKGNMADMDHDIWPMSLVDANAGVW 252
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 155 FSPRDWQRLNNNGWFRPKRLNFKLNIQVKEVTND---GVTTIANNLSTVQVPSDS 210
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 253 LQPSDWQYICNTMGQLNLVSLDQEIFNVVLKTVTEQDGGQAIKTYNNDLTAQMMVAVDS 312
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 211 EYQLPYVLSAHQCLPPPADVFEMIPQYQ-----LTNNGSQAV-----G 252
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 313 NNILPYPPAASMETLGFYPMKPTTASPRYFCVDRDLSTVENQGTVEHNVMPKPG 372
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 253 RSFYCLEFPPS-OMLRGNNF-TFSYFEEVPPHSSVAHSQSLDRLMN--PLIDYLYX 308
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 373 IPQFTIENTQOITLRTGDEFATGTYFDINSV--KLTHTWQNNRQGLGQPPPLSTF--- 427
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 309 LNRTONSGSAONKDLFSRGSPPAGMSVQPKNL-----PG 344
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 428 -PEADTLAGT-----LTAQGRHGTQMGVNWVSEAIRTRPAQVGFQPHNDPEASRAG 480
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 345 PCYVQGVSKTKTNNNNNFTWTGASKYNLNGRESIINPG-----TAMASHKODE 394
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 481 P-FAPKVPADITQGVDEKANGSVRYSGKHGEMASHGAPRRYTWDETSPSGSDTK 539
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 395 DKFF-----PMGVMIFGKESAGASANTALDNVMTDEEIKATNPVATERFGTAV 445
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 540 DGFQASAPLVVPPPLNGI-----LTNNPISGTKN-----DI 570
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 446 NFQGS-SSTDPAIDVHMGALPGKWODRDVYLQGPIMAF--IPTHGHHPSPLMG 500
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 571 HFSVVFPSYGLTAFSSH-----PSPVYPOGQIWDKELDEHKPRLHITAPPV-- 617
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 501 FGLKNPPQILIKNTP--VPANPPAPBSATKFAPIIYQSGVSEIEMELQKNSKRW 558
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 618 -CKNNAPQGMRLRGLPILTDQYDPNGATLSRIYVIGTFWKGKLTMPAKLRA---WTTW 672
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 559 NPEVQYTSNVAKSANVDFTVNNGLY--TEPRIGT-----RYLTPPL 599
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 673 NPVQVSAE-----DNGNSYMSVTKMLPTATGNNQSVLITRPV 711
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 11
 VCPVIF
 coat protein VP1 - feline panleukopenia virus

N:Contains: coat protein VP2
 C:Species: feline panleukopenia virus, FPLV
 C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
 C/Accession: A03701
 R/Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winston, S.; Hahn, W.
 J. Virol. 55, 574-587, 1985
 A/Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parv
 A/Reference number: A03697; MUID:85265017; PMID:2991581
 A/Accession: A03701
 A/Molecule type: DNA
 A/Residues: 1-727 <CAR>
 A/Cross-references: EMBL:M10824; NID:G333474; PIDN:AAA7161.1; PID:G333476
 C/Genetics:
 A/Intons: 11/1
 C/Superfamily: parvovirus coat protein
 C/Keywords: coat protein

F.144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.6%; Score 216; DB 1; Length 727;
 Best Local Similarity 20.2%; Pred. No. 6.4e-07;
 Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

```

QY 7 RVEQSPQEPDSSSGIGTKGQAPAKRLNFGQTGDESVV---DQPLGEPPA--TPAA 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 110 KPTKRSKPPHIFINLAK-----KKKAGAGQYKRDQNAQMSDGAQVQDGGQAVANERA 163
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 VQPTTASGGGAPADNNEGADVGNASGWH-----CDSTWLDGRVITSTRWALP 113
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 164 TGSNGSGGGG-----GGSGGVGISTGFNNQTEFKLENGV--ETANSSRLVHLN 215
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 114 TNNHLYKQI-----SSASTGASNDNHYFGYSTPMGFYDNRPHCHSPRDWQRLNN 166
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 216 MPESENTKRVVANNMKTAVKGMALDDHVOQVLTWMSLVDANAKGWRNPGMQLIVNT 275
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 167 NMGFRPRRLNFKLNIQVKEVT--TNDGVTIANNLSTVQVPSDSYQLPYVLSAHQ 223
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 276 MSELHLVSPQEIFNVVLKTVSBSATQPTKYVNNDLTASIMVALDSNNTMPTPAMRS 335
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 224 GCL-----PPPA-----DVFMIPOYGYLTNNGSQAVRS-----SFYCL 259
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 336 ETLGFYPMKPTTIPWRKYFQWDRTLPSH-----TGSPTNTYHGTDPDDVQFYTI 389
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 260 B-YFSPQMLRTGNF-TFSYFEEVPPHSSVAHSQSLDRLMN--PLIDYLYXNRTQNG 315
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 390 ENSVPHILRTGDEFATGTFPDKP--CRHTTWQNNRQGLGQPPPLSTF-----PLNSLPQS 440
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 316 SSGAONKDLFSRGSPPAGMS-----VQPKN-WLPQCYRQORVSK----- 354
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 441 EGATNFGDIDQVQDKRGVLTQMGNTDYTEATIMRAEVGSAFYVSFEASTGQPFKPI 500
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 355 -----TKTDNNNSNTWTGASKY---NNGRESIINPG-----AASHKDDDKFRPMS 401
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 501 AAGRGAGQIDENAA--DQPRYARQHQGQTTTGTGPERFYIAHDT----- 549
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 402 GWMI FGKESAGASNTALD-NVMTDEEIKATNPVATERFGTAVANFOSSSTDPATGVA 460
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 550 -----GRYPAGDMIQINIRPLPTNDNVLPIDPIG-----GKGINV--TININTYGP 598
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 461 AMGALPGMWODRDVYLQGPIMAFIPTHGHHPSPLMGFGLK-----NPP 507
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 599 ALNNVP-----PVYPNGQIWDKEDTD-----LKPRLHVANAPVCONNCP 638
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 508 POLINKTPVPAN--PPAPBSATKFAPIIYQSGVSEIEMELQKNSKRWNEVQY 564
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 639 GOLFVAVAPVLTNEYDPAASANSR---IYVSDPFWMKGLVFKAKLRASHTWNPYQQM 694
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 565 TSNVAKSANVDFTVNNGLYTEPRIG 591
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 695 SIN-----VDNQFNYL-PNNIG 710
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 12
 VCPVIM

coat protein VP1 - minute virus of mice (strain WVM1)

C:Species: minute virus of mice, murine parvovirus
 C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 28-Jul-2000
 C/Accession: B23008; B29510
 R/Sahli, R.; McMaster, G.K.; Hirt, B.
 Nucleic Acids Res. 13, 3617-3633, 1985
 A/Title: DNA sequence comparison between two tissue-specific variants of the autonomous I
 A/Reference number: A23008; MUID:85242059; PMID:3855242
 A/Accession: B23008
 A/Molecule type: DNA
 A/Residues: 1-718 <SAH>
 A/Cross-references: EMBL:X02481; NID:960918; PIDN:CA846507.1; PID:95419928
 R/Asell, C.R.; Gardiner, E.M.; Tattersall, P.
 J. Virol. 57, 656-669, 1986
 A/Title: DNA sequence of the lymphotopic variant of minute virus of mice, WVM(1), and c
 A/Reference number: A29510; MUID:86115415; PMID:3502703

A:Accession: B29510
 A:Molecule type: DNA
 A:Residues: 1-143, 'A', 145-718 <AST>
 A:Cross-references: EMBL:M2032
 C:Superfamily: parvovirus coat protein
 C:Keywords: coat protein

Query Match 6.6%; Score 214.5; DB 1; Length 718;
 Best Local Similarity 21.0%; Pred. No. 8e-07;
 Matches 142; Conservative 89; Mismatches 273; Indels 171; Gaps 30;

```

QY PVEQSPQEPDSSSGIGKTG-----QQPAKRLNFG---QTGSESVDPDPLG 52
DB PKLATSEF-GTSGVSAAGKTRTPAYIFINQARAKKLTSSAAQSSQTMSTGTPDG 141
QY 53 -----EPPTPAVGTTPMAGGAGMADNNEGADVGNAGNWHGDS--TWLGD---R 101
DB 142 GNGVSAARVERAADGG--GSGGG-----GSGGGGVGVTGTSYDQTHYRFLGDMWE 193
QY 102 VITTSRTMALPTYNHLYKQI-----SSASTGASNDNHYFGYSTPMGYFPDPRPHCH 154
DB 194 ITALATITLVHNPKSNTYCRIRVHNTTDSVKGNNAKDADHEQITPMSLVDAANAGW 253
QY 155 FSPRDWQRLINNNGFPRKRLNFKLQVKEVTTND---GVTTANLSTVQVPSDS 210
DB 254 LQPSDMQYICNTMSQNLVSLDQEIFNVVLKTVTEQDSGGQAIKYNNDLTACMVAVDS 313
QY 211 EYQLPVYLSAGAGCCLPPPADVPMIPQYV-----LTLNN-----GSAV 251
DB 314 NNLPYTPPANSKMETLGFYPMKPTIASPYRYFCVDRDLSTYENOBGTEHNVMGPKG 373
QY 252 GRSSFCLEYEFP--QMLRTGNF--TFSTPEEVPFHSVYASQSLDRLM--PLIDQYLY 307
DB 374 MNSQFTIENTQOITLRTGDEPATGTYTFDINPV--KLTHWTQTRQLGQPLSTF-- 429
QY 308 YINRTONQSGAONKDLLFSRGPAGMSVOPKNWL-----P 343
DB 430 --PEADTDAGT-----LTAQSGRHATQMEVNMVSAIRTPAQVFCQPHNDFEASRA 481
QY 344 GPCYRQORVSKITDNNNSNFTWTGASKNLNGRESIINPGTAMASKDEDEK----- 397
DB 482 GP-FAAPKVPADVYQGVDRBANGSVRSYKQHEENNAAGPAPERWTMBETFGSGRDT 540
QY 398 ---FPMGWNIFGKESAGSANTLNDNMITDEBEIKATNPVATERFETAIVNFOS--SST 452
DB 541 RDGTQSAPLVPPPLGLITNA-----NPLGTN---DIHFSNPNSTY 581
QY 453 DPATGDVHANGALFGWVQDRDVTYLGSPIMAK---IPHTDGHFSPFLMGFGGLKNPPQ 509
DB 582 GPLTAFSH-----PSPVYPOQIMDKELDLHKRRLHITAFV---CKNNAPOQ 627
QY 510 ILIKNTP--VPANPFAFSATKFASTFYQSTGOVSVEIEMELQKNSKRNPEVQYTSN 567
DB 628 MLVRLGPNLTDOYDPCGATLSRIYTGTFPMWKLTRAKLRA---NTTWNPVYQ----- 679
QY 568 YAKSANVDFTVNNNG 582
DB 680 -----VSVEDNG 686

```

RESULT 13

VCPVME

coat protein VP1 - mink enteritis virus (strain Abashiri)

N:Contains: coat protein VP2

C:Species: mink enteritis virus, MEV

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 12-Apr-1996

C:Accession: B38350

R:Kariatsunari, T.; Horinouchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinag

J. Gen. Virol. 72, 867-875, 1991

A:Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the

A:Reference number: A38350; MUID:91202123; PMID:2016597

A:Accession: B38350

A:Molecule type: DNA

A:Residues: 1-722 <MAR>
 A:Cross-references: GB:D00765
 C:Superfamily: parvovirus coat protein
 C:Keywords: coat protein
 F:139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.6%; Score 214; DB 1; Length 722;
 Best Local Similarity 20.2%; Pred. No. 8.7e-07;
 Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

```

QY 7 RPYEQSPQEPDSSSGIGKTGQPAKRLNFGQTGSESV-----DPQIGPPA--TPAA 60
DB 105 KPTKRSKPPPHIFINLAK-----KKKAGAGVYKRDNLAPMSDGAQVDPGGPAPVANEBA 158
QY 61 VGPPTMASGGAAPMADNNEGADVGNAGNWH-----CDSTWGLDRVITTSRTMALP 113
DB 159 TSGNGSGGGG-----GGSGGVGISTGTFNNQTEFKLENGWV--ETANSRLVHLN 210
QY 114 TYNNHLYKQI-----SSASTGASNDNHYFGYSTPMGYFPDPRPHCHSPRDWQRLINN 166
DB 211 MPSENYKAVVANNMOKTAVKGMALDDTHVQLVTFMSLVDAANAGWENFPGMOLIVMT 270
QY 167 NMGFPRKRLNFKLQVKEVTTNDGVTTIANNLSTVQVSDSEYQLPYVLSAQH 223
DB 271 MSELHVSFEQEIFNVVLKTVSESATQPTKYVNNDLTASLWALDSNNTMPTPAMRS 330
QY 224 GCL-----PPPPA-----DVEMIPQYVITLNGSQAVERS-----SFYCL 259
DB 331 ETLAGYPMKPTLPTPMRYVPOMDRTLPSH-----TGSCTPTNLYHGTDPDVQFYTI 384
QY 260 E-YFSPQMLRTGNF--TFSTPEEVPFHSVYASQSLDRLM--PLIDQYLYINRTONO 315
DB 385 ENSVPHLARTGDEPATGTFPFDCR--CRLTHWTQTRALGLP-----FLNSLPQS 435
QY 316 SSGAONKDLLFSRGPAGMS-----VQKN-MLPQCYRQORVSK----- 354
DB 436 EGATNFGDIGNQDKRGVTOGMENTDYTEATIMRAEVGYSAPYSPFASTGSPKPTDI 495
QY 355 -----TKTDNNNSNTWTGASKY---NNGRESIINPGT---ANASHKDEDKFFPMS 401
DB 496 AAGRGAAQDENQAA--DQDPYAFGRQHQQLTTTGTEPERFYIAHQDT----- 544
QY 402 GWMIFGKESASNTALD--NVMITDEBEIKATNPVATERFETAIVNFOSSTPATGDVA 460
DB 545 -----GRYAGDMIQININNLPTNDNVLLPDPDIG---GKTGINY--TNIFNTYGPLT 593
QY 461 AMGALPGMWQDRDVTYLGSPIMAKIPHTDGHFSPFLMGFGGLK-----NPP 507
DB 594 ALNNVP-----PVPYPCQIMDKEDTD-----LKPRHLVNAAPFVCONNCP 633
QY 508 POLIKNTVPAN--PPAFSATKFASTFYQSTGOVSVEIEMELQKNSKRNPEVQY 564
DB 634 GOLFTVAAPLNEYPDASANSR---ITVYDFPMWKGKLVKAKLRASHWTNPIQOM 689
QY 565 TSNYAKSANVDFTVNNNGLYTERPPIG 591
DB 690 SIN-----VDNQPNYL--PNNIG 705

```

RESULT 14

VCPVFP

coat protein VP1 - feline panleukopenia virus (strain 193)

N:Contains: coat protein VP2

C:Species: feline panleukopenia virus, FPV

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 12-Apr-1996

C:Accession: B36608

R:Marlyn, J.C.; Davidson, B.E.; Studdert, M.J.

J. Gen. Virol. 71, 2747-2753, 1990

A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvo

A:Reference number: A36608; MUID:91073139; PMID:2174965

A:Accession: B36608

A:Molecule type: DNA

A:Residues: 1-727 <MAR>

A:Cross-references: GB:X55115
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
F:144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.4%; Score 209; DB 1; Length 727;
Best Local Similarity 20.1%; Pred. No. 1.9e-06;
Matches 138; Conservative 101; Mismatches 260; Indels 188; Gaps 34;

```
QY 7 RVPQSPQEPDSSSGIGCTGQAPAKKRLNFGQDSESV-----DPOLEGPPA--TPAA 60
DB 110 KPTKRSKPPHIFINLAK-----KKKAGAGYKRDNLAPMSDGAVOPEGQPAVNERA 163
QY 61 VGPPTMASGGGAPMADNNEGADGVGNASGNW-----CDSTWLGDRVITSTRMALP 113
DB 164 TGSNGSGGG-----GGSGGAGVIGSTGTFTNNQTEKFLKGV--EITANSRLVHLN 215
QY 114 TYNHLYKQI-----SSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDMQRLINN 166
DB 216 MPSESNYKRVVNNMDKTAIVGNMALLDIIHQIVTPEMSLDANAMGWFFNGDMQLIVNT 275
QY 167 NWGRRPKRLNFKLFINQKVEY---TNDGVTIANNLSTVQVPSDSYQLPYLGSAAHQ 223
DB 276 MSELHLVSFEQEIFNVYLKTVSESATQPTKYVNNDLTASLWALDSNNTMPTPAMRS 335
QY 224 GCL-----PPPPA-----DVFMIPQGYTLNNGSAVGRS-----SFYCL 259
DB 336 ETLGFPYKPKTITTPMKRYFQMDRLIIPSH-----TGTSGTPTNHYHGTPDDVOGYTTI 389
QY 260 E-YFPQMLRTGNF--TFSYTFEEVPHSVAHSQSLDRLN--PLIDQVLYLNRTQNG 315
DB 390 ENSVPVHLRTGDEFATGTFEFDCKP--CRLTHQTNRALGLP-----FLNSLPQS 440
QY 316 SGSAQNKDLLFSRGSAGMS-----VQPKN-WLPGCYQQRVSK----- 354
DB 441 EGATNFGIIGVQODRRGVTOGNTDYTEATIMPAEVGSAPYSFEASTQGPFTPI 500
QY 355 -----TKDNNNSNFTWTSKASY---NINGRESIINPT---AMASHKDEKFFPMS 401
DB 501 AAGRGAGQTDENQA---DGDPRYAFGRHQOKTTTGETPERFYIAHQDT----- 549
QY 402 GWMITFGESAGASNTALD--NMTIDDEEIKATNPATERFGTVAVNFQSSSTDPATGDVH 460
DB 550 -----GRYEGEDWIONINFNLPTNDVLLPTDPIG---GKTGINY--TNIFNTYGPIT 598
QY 461 AMGALPGWVQDRDYLQPIWAKIPHTDGHFHSPLMGFGK-----NPP 507
DB 599 ALANVP-----PYVNGQIMDKEDTD-----LKPRLHVAAPVCCNNCP 638
QY 508 PQILIKNTPVAN---PPAEFSATKFAFTQYSTGVSEIEMELQKNSKNNPEVOY 564
DB 639 GQLPVKAPNLTNEYDPDASANMSR---IVTSDPFWMKKGLVFKAKLRASHTNPPIQOM 694
QY 565 TSNYAKSANVDFTVDNNGLYTEPRPIG 591
DB 695 SIN-----VDNQFNIV--PNNIG 710
```

RESULT 15

VCpVCD
coat protein VP1 - canine parvovirus (strain CPV-2)
N/Contains: coat protein VP2
C/Species: canine parvovirus CPV
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C/Accession: A31163
R/Parish, C.R.; Aquadro, C.F.; Carmichael, L.E.
Virology 166, 293-307, 1988
A/Title: Canine host range and a specific epitope map along with variant sequences in th
A/Reference number: A31163; MUID:89020796; PMID:3176341
A/Accession: A31163
A/Molecule type: DNA
A/Residues: 1-737 <PAR>
A/Cross-references: EMBL:M23255; NID:G333467; PIDN:AAA47158.1; PID:G333468

C/Genetics:
A:Initons: 26/3
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
F:584-737/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.4%; Score 208; DB 1; Length 737;
Best Local Similarity 19.8%; Pred. No. 2.3e-06;
Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;

```
QY 7 RVPQSPQEPDSSSGIGCTGQAPAKKRLNFGQDSESV-----DPOLEGPPA--TPAA 60
DB 120 KPTKRSKPPHIFINLAK-----KKKAGAGYKRDNLAPMSDGAVOPEGQPAVNERA 173
QY 61 VGPPTMASGGGAPMADNNEGADGVGNASGNW-----CDSTWLGDRVITSTRMALP 113
DB 174 TGSNGSGGG-----GGSGGAGVIGSTGTFTNNQTEKFLKGV--EITANSRLVHLN 225
QY 114 TYNHLYKQI-----SSASTGASNDNHYFGYSTPWGYFDNRFCHFSPRDMQRLINN 166
DB 226 MPSESNYKRVVNNMDKTAIVGNMALLDIIHQIVTPEMSLDANAMGWFFNGDMQLIVNT 285
QY 167 NWGRRPKRLNFKLFINQKVEY---TNDGVTIANNLSTVQVPSDSYQLPYLGSAAHQ 223
DB 286 MSELHLVSFEQEIFNVYLKTVSESATQPTKYVNNDLTASLWALDSNNTMPTPAMRS 345
QY 224 GCL-----PPPPA-----DVFMIPQGYTLNNGSAVGRS-----SFYCL 259
DB 346 ETLGFPYKPKTITTPMKRYFQMDRLIIPSH-----TGTSGTPTNHYHGTPDDVOGYTTI 399
QY 260 E-YFPQMLRTGNF--TFSYTFEEVPHSVAHSQSLDRLN--PLIDQVLYLNRTQNG 315
DB 400 ENSVPVHLRTGDEFATGTFEFDCKP--CRLTHQTNRALGLP-----FLNSLPQS 450
QY 316 SGSAQNKDLLFSRGSAGMSVQPKMLPGCYQQRVSKTKDNNNSN-----TW 366
DB 451 EGATNFGDI-----GV-----QDQKRRGVTOGNTDYTEATIMPAE 488
QY 367 TGASKYLNGRESIINP-----GTAMASHKDEKFFPMSGVMIFGESAGASNTALD-- 419
DB 489 VGSAPYPSFEASTQGPFTPIAAGGAGQTDENQAADGNPRAFRGRHQOKTTTGETP 548
QY 420 -----NMTIDDEEIKATNPATERFGTVAVNFQSSSTDP 454
DB 549 ERFYIAHQDTGRYPRGDMIONINFNLPTNDVLLPTDPIG---GKTGINY--TNIFN 602
QY 455 ATGDVAMGALPGWVQDRDYLQPIWAKIPHTDGHFHSPLMGFGK----- 504
DB 603 TYGPLETALANVP-----PYVNGQIMDKEDTD-----LKPRLHVAAPV 642
QY 505 ---NPPQILIKNTPVAN---PPAEFSATKFAFTQYSTGVSEIEMELQKNSKNNPEVOY 558
DB 643 CQNNCGQLFVKAPNLTNEYDPDASANMSR---IVTSDPFWMKKGLVFKAKLRASHTN 698
QY 559 NPEVQTSNYSKANSVDFTVDNNGLYTEPRPIG 591
DB 699 NPICQMSIN-----VDNQFNIV--PSNIG 720
```

Search completed: July 17, 2003, 18:35:46
Job time: 23.473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 11.8582 Seconds

(Without alignments)
2095.115 Million cell updates/sec

Title: US-09-807-802a-15

Sequence: 1 TAGGKRPVQSPQEPDSSS.....NNGLYTEPRPIGTRVLRPL 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	67.0	504	COA3_AAV2	P03135 adeno-assoc
2	556	17.1	673	COAT_PAVBO	P07297 bovine parv
3	495	15.2	781	COAT_PAVHB	P07299 human parv
4	259.5	8.0	749	COA2_PAVBN	P23964 porcine par
5	254.5	7.8	729	COA1_PAVP9	P33484 porcine par
6	254.5	7.8	729	COA1_PAVN	P18546 porcine par
7	252.5	7.5	729	COA1_PAVK	P2501 porcine par
8	244	7.2	587	COAT_PAVH	P03136 hamster par
9	234	7.2	587	COAT_PAVJ3	P63310 parvovirus
10	216	6.6	716	COAT_MMIV	P03137 murine min
11	216	6.6	727	COAT_FPV	P04864 feline min
12	214.5	6.6	718	COAT_MMIM	P07302 murine min
13	214	6.6	722	COAT_MEVA	P07437 mink enteri
14	209	6.4	727	COAT_PAV19	P24840 feline panl
15	208	6.4	727	COAT_PAVCB	P11213 canine panl
16	208	6.4	737	COAT_PAVCD	P17455 canine parv
17	204	6.3	748	COAT_PAVCN	P12930 canine parv
18	195	6.0	584	COAT_PAVC7	P30129 canine parv
19	193	5.9	722	COAT_PAVC	P04863 canine parv
20	186.5	4.7	647	COAT_PAVG	P24029 aleutian mi
21	143.5	4.4	648	COAT_ADVG	P61575 mus musculu
22	123	3.8	648	WHN_MOUSE	O15153 homo sapien
23	123	3.8	648	WHN_HUMAN	O15153 homo sapien
24	121	3.7	1142	ENAM_PIG	O97939 sus scrofa
25	121	3.7	880	SYV_BACST	P11931 bacillus st
26	119	3.7	1113	N116_YEAST	Q02630 saccharomyc
27	116.5	3.6	667	ITF2_HUMAN	P15884 homo sapien
28	110.5	3.4	655	ITF2_KLEBN	P15881 canis fami
29	110.5	3.4	655	CDGT_KLEBN	P08704 klebsiella
30	109.5	3.4	1742	GUNA_CALSA	P22534 caldocellum
31	109.5	3.4	531	HEXB_PIG	Q29548 sus scrofa
32	109	3.4	670	ITF2_MOUSE	O60722 mus musculu
33	108.5	3.3	1186	CEAA_BACTS	Q45710 bacillus th
33	108.5	3.3	860	AREA_PENRO	O13508 penicillium

34	108.5	3.3	1379	1	YFP2_SCHPO	O14066 schizosacch
35	108.5	3.3	1849	1	IGA4_HAEIN	P45386 haemophilus
36	107	3.3	2493	1	CYAA_USTMA	P49606 usellaga ma
37	106.5	3.3	1070	1	Y355_HUMAN	O15063 homo sapien
38	106	3.3	1271	1	Y338_MYCGE	P47580 mycoplasma
39	105.5	3.2	559	1	HNEP_PIG	O03365 sus scrofa
40	105	3.2	1396	1	ITAA_DROME	P12080 drosophila
41	105	3.2	1656	1	OMPE_RICUA	O06533 r outer mem
42	104.5	3.2	1717	1	CLAI_ARATH	O38854 arabidopsis
43	104.5	3.2	1122	1	ADP1_MYCGA	Q49379 mycoplasma
44	104	3.2	395	1	MAFI_YEAST	P41910 saccharomyc
45	103.5	3.2	1196	1	AMVB_PABPO	P21543 paenibacill

ALIGNMENTS

RESULT 1	COA3_AAV2	STANDARD;	PRT;	504 AA.
ID	COA3_AAV2			
AC	P03135;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-NOV-1991 (Rel. 20, Last annotation update)			
DE	Probable coat protein 3.			
OS	Adeno-associated virus 2 (AAV2).			
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
OX	NCBI_TaxID=10804;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=83164299; Pubmed=6300419;			
RA	Srivastava A., Luby E.W., Berns K.I.;			
RT	"Nucleotide sequence and organization of the adeno-associated virus 2			
RL	J. Virol. 45:555-564 (1983).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL, outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J01901; AAA42376.1; -.			
DR	PIR; A03698; VCPV3A.			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
KW	Coat protein.			
SQ	SEQUENCE 504 AA; 56366 MW; 758999B017052B6B2 CRC64;			
Query Match	67.0%; Score 2177.5; DB 1; Length 504;			
Best Local Similarity	82.5%; Pred. No. 3.4e-137;			
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;				
QY	66 MASGGAPADANNNEGADVGNAAGNHCSTWIGDRVITTSRTTMAIPYNNHLYQIIS 125			
DB	1 MATGSGAPADANNNEGADVGNSGNWCHDSTWIGDRVITTSRTTMAIPYNNHLYQIIS 60			
QY	126 ASTGASNDNHYPGYSTPMGTFDNRPHCHSPSPDMQRLINNNGFPRKRLNPLFNIOVK 185			
DB	61 QS-GASNDNHYPGYSTPMGTFDNRPHCHSPSPDMQRLINNNGFPRKRLNPLFNIOVK 119			
QY	186 EYTTNDGVTTIANNLTSTVQVFSDSSEYQLPYVLGSAHQGLPFPADVPMIPGYGITTN 245			
DB	120 EYVQNGTTIANNLTSTVQVFTDSSEYQLPYVLGSAHQGLPFPADVPMIPGYGITTN 179			
QY	246 NSGQAGRSFYCLFEPSCMLRTGNPFSSYFEEVPPHSSYFAHSQSLDRLNMPILDOY 305			
DB	180 NSQAGRSFYCLFEPSCMLRTGNPFSSYFEEVPPHSSYFAHSQSLDRLNMPILDOY 239			
QY	306 LVTIANTQNGSAGKQKDLIFSRGSPAGMSVQPKNMLPGFCYRQQRVSXTKTNNNSNFT 365			

```

Db 240 LYLSTNTPTSGTTQSLRQPSQAGASIDRDSRWMLPGFCYRQGRVSKTSADNNSEYS 299
Qy 366 WTGSKTNLNGRETIINPTGMASHKDEDEKFFPMSCGMITGKASAGASNTALDNWITD 425
Db 300 WTGATKYNLNGRDLNVP--AMASHKDEDEKFFPMSCGMITGKASAGASNTALDNWITD 357
Qy 426 EEEKATNPVATEFGTVAVNFQSSSTDPATGVHANGALPGWQDRDYLQSPINAKI 485
Db 358 EEEIGTNPVATEFGTVAVNFQSSSTDPATGVHANGALPGWQDRDYLQSPINAKI 417
Qy 486 PHTDGHFHPSPLMGFGFLKKNPPQILIKNTVPANPAEFSATKFSFTQYSTG 540
Db 418 PHTDGHFHPSPLMGFGFLKKNPPQILIKNTVPANPAEFSATKFSFTQYSTG 472

RESULT 2
ID COAT PAVBO STANDARD; PRT; 673 AA.
AC P07257; 084374;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP2 [Contains: Coat protein VP3].
OS Bovine parvovirus (BPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=10784;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87061184; PubMed=3783814;
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,
Bates R.C.;
RT "Complete nucleotide sequence and genome organization of bovine
parvovirus."
RL J. Virol. 60:1085-1097(1986).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; M14363; AAB59848.1; -
DR EMBL; M14363; AAB59848.1; -
DR PIR; A26104; VCPVB5.
DR HSSP; P30129; ADPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 673 COAT PROTEIN VP2.
FT CHAIN 138 673 COAT PROTEIN VP3.
FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 163 187 GLY-RICH.
SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 17.1%; Score 556; DB 1; Length 673;
Best Local Similarity 25.0%; Pred. No. 1.5e-29;
Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;
Qy 1 TAPGKKRPVEQSPQEPDSSSGIGTGQOPAKRLNFGQTGSEVPPDPLGEP--PATP 58
Db 100 TSKGGBALKKRLYFASRNKAKRANREPASTNSQNMVEVSNIPENDEAGNCEIATLR 159
Qy 59 AAVGPTMASGGGAPMADNMGAGVGNAGNMHDSFWLDRVITTSRTWALPTNNH 118
Db 160 SVVSGSVGGGG-----RGSGGVGISTGWMGGTITFSNIVITKTRPFDIDIKGH 211
Qy 119 LYKOISSASTGASNDNHFGYSTPWGYDFNRFCHGSPRDMORLINNMGFRKRLNFK 178

```

```

Db 212 LYKS-EVLANTGDIAHROY-ALTTPWGYFNENQVSHFSPMDWQHLYNDVRRFPKMIYR 269
Qy 179 LFNIVQKEVTTNDQVTTIAN--LITVQVFSDEYQLPYVLSAAGCLPPFADVPMIP 237
Db 270 VVNLQIKQIMTIDAMGVVYNNDLTACGHI-PCDDHRYPYQHPWDQCMELFNLSIWELP 329
Qy 238 QYGYVLT-----NNGSQAVGRS-----SFCLYEPPSQMLRTGNNPTFSYTEEYVPHS 286
Db 330 QYVITPAPISVDNNTNTVVEHLKGVPLYLENSDHEVLRNG----- 373
Qy 287 SYAHQSGLDELNPLIDYLYYLNRQN-----QSSGAQNKDLFSGRGP----- 331
Db 374 -----RIYRIYQLWRLBMDRQOHNIQHASDDVQSGQKNNLLIQRTPQNKORF 424
Qy 332 AGMSVQPKWMLPQPCRCQQRQVSKTKTDNNSNTFTWAGSKYNNNGRETIINPTGMASHK 391
Db 425 QNALRLTSMWMSGP-----GIARGTNNATLQTSAGALVTMT 462
Qy 392 DDEDKFPFMSGVM-----IFGKE-----SAGASNTALDNWITDE 426
Db 463 NGAD-----VSGVAVRVGISTDPTIYGQCPESDLRLRYSAAGQONPLEN----- 512
Qy 427 EEIKATNPVATERFGTVAVNFQSSSTDPATGVHANGALPGWQDRDYLQSPINAKIP 486
Db 513 -----AARHFTREARFKLITGSNGADGDKEMWMLPNQWDSAPISRYPDIWKVP 564
Qy 487 HTDGHFHPSPLMGFGFLKKNPPQILIK--NTVPANPAEFSATKFSFTQYSTQGVSV 544
Db 565 RVNRKTLTLDQDSIPMSHPPGTFILKARIIPVGGND-----SPLNIVYTGQVSC 615
Qy 545 EIEMLQKNSKRNPNPEVQYTSNYAKSAND-PTVNNNGLY 584
Db 616 EVWVEVKEKGTQWRREYMH5---ATNMSYDATINNAGY 653

RESULT 3
ID COAT PAVHB STANDARD; PRT; 781 AA.
AC P07259;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Probable coat protein VP1.
OS Human parvovirus B19.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
NCBI_TaxID=10798;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Isolate AU;
RX MEDLINE=86200451; PubMed=3701931;
RA Shade R.O., Blundell M.C., Cotmore S.F., Tatefaleli P., Astrell C.R.;
RT "Nucleotide sequence and genome organization of human parvovirus B19
isolated from the serum of a child during aplastic crisis."
RL J. Virol. 58:921-936(1986).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; M1178; AAB68867.1; -
DR PIR; A24299; VCPV19.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;
 Query Match 15.2%; Score 495; DB 1; Length 781;
 Best Local Similarity 27.3%; Pred. No. 2e-25;
 Matches 148; Conservative 84; Mismatches 237; Indels 54; Gaps 16;

QY 51 LGEPPTAAVGTPTNAGGGAPMADNNEGADVGNAGNMGHCDSTWLGDRVITTTSTRTW 110
 DB 213 LPEVPVAPNASEKPSMTSVNSAB-ASTGAGGGGNSVKSMMWSGATPSANVCTPSROF 271
 QY 111 ALPTNNHLYKQISSAGTGSND-----NHTFGYSTPMGYPDPFNHCHFSPRDW 160
 DB 272 LIEYDEPHHYKVPSPASSCHNAGSKRAKVCITSPINGSTPMRYLDPNALNFFSPLER 331
 QY 161 QRLINNNWGRPRKLNKLFNIOVEYT--TNDGVTTIANLSTVOVSDFEYQAPYVL 218
 DB 332 QHLIENYGSIAFDALTTTISEINAKVDYDKTGGGV-OYDSTTGRCLMAYDHEKIPYVL 390
 QY 219 GSAHOGCLPPPADVFPMIPQYGYLTLLN-GSQAVG-----RSSFYCLEYFPPSQMLR 268
 DB 391 GGGQDTLAPELPIWVYPPQYAVLTVDVMTQGISGSKLASBESAFYVLEHSSPQLLG 450
 QY 269 TGNNTFTSYTPEEVPFHSYAHQSGLDRMLNPLIDQYLYLNTQNGSAGAKNDLFSR 328
 DB 451 TGGTASMSYKFPVPPEPNTLEGCSQHFEYEMYNPL---YGSRLGVPDILGDPKFRSL---- 503
 QY 329 GSPAGMSVOPKMLPGPCYQGRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPG-TAM 387
 DB 504 -THEDHAIQONTPMPGLVNSVSTKEGSSNTGAKALTELSTGTSONTIRISLRPGVSQ 562
 QY 388 ASHDEDEKFPMSGVMIPKESAGASNTALDNV-----MITDEEIKATVPVATERP 440
 DB 563 PYHMDTDKVTGIMNISHGQTYG--NAEDKRYQGVGRFPNEKQQLQGLMNHVY 619
 QY 441 GTAAVNFQSSSTPAPADVAMGALPGMVDQDQVYIQGPIWAKIPIPTHGHPS-PLMG 499
 DB 620 -----FPNKGTOQYTDQIE-RPLMVGSVNRRRLHESQJMSKIPILDISFTQFALG 672
 QY 500 GFGKAPPPQILIKNTVPANPAEFSATKFAPIITQSTGOVSVEIEMEL-OKENSKRM 558
 DB 673 GMLGHDPPIPLK--ILPQSGPIGKIKSGITTLVQYAVGIMVTWTFKLGPRKATGRM 730
 QY 559 NPE 561
 DB 731 NPQ 733

RESULT 4
 COA2 PAVPN STANDARD; PRT; 749 AA.
 ID COA2 PAVPN STANDARD; PRT; 749 AA.
 AC P2364;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2] (Version 2).
 OS Porcine parvovirus (strain NADL-2) (PPV).
 OS Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OC NCBI_TaxID=10797;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90010964; PubMed=2794971;
 RA Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;
 RT "Porcine parvovirus: DNA sequence and genome organization.";
 RL J. Gen. Virol. 70:2541-2553(1989).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC -1- VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D00623; BAA0502.1; -
 DR PIR; B3302; VCPVP.
 DR HSP; P30129; ADPV.
 DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 749
 FT DOMAIN 195 210
 FT CARBOHYD 192 192
 FT CARBOHYD 218 218
 FT CARBOHYD 302 302
 FT CARBOHYD 350 350
 FT CARBOHYD 453 453
 FT CARBOHYD 491 491
 FT CARBOHYD 593 593
 FT CARBOHYD 624 624
 FT CARBOHYD 671 671
 SQ SEQUENCE 749 AA; 82872 MW; 6CA13CB97BE93418 CRC64;
 Query Match 8.0%; Score 259.5; DB 1; Length 749;
 Best Local Similarity 23.1%; Pred. No. 7.9e-10;
 Matches 155; Conservative 92; Mismatches 270; Indels 153; Gaps 35;

QY 9 VEGSPQEPDSSSGIGK-----TGOQPAKRLNFGQTDGSDSEVPDPQIGBPAP 58
 DB 129 VRSPRKFGSKRPGRPAPRHIFIMLAKKAKGTNTNSMSSENVQHNPIIN-AAVE 186
 QY 59 -AAVPTTMAAGGAMADNNEGADVGNAGNMGH--CSTWLGD--RVITTSITWML 112
 DB 187 LSAITNESGGGGGGG---GGRGAGGVSTGSEFNQTFQYIGKELVITAHASRLIH 242
 QY 113 PTYNHLYKQI-----SSASTGAS-NDNHYFGSTPMGYPDPFNHCHFSPRDWQRLNN 166
 DB 243 NMPEHTTYRIHVLNBSGASAGMVDQDQVYIQTWMSLIDANAGWENRPAQMLISNN 302
 QY 167 NMGFRKRLNFKLNIQVEYV--TNDGVTTIANLSTVOVSDFEYQAPYVLGSAHQ 223
 DB 303 MTEINLVSEQALFNVLKLTITESATSPPTKLYNNDLFTSLMVALDNTNLTPTAPARS 362
 QY 224 GCLPPPADVFPMIPQYGYL-----TLNNGSAVGRS-----SFCLE-YRP 263
 DB 363 ETLGFPYMLPTKPYRYVLSCLRNLNPPTYGSOQITDSIQTGASHDIMFYTIENAVP 422
 QY 264 SOWLRGTGNP-TPSYTPEEVPFHSYAHQSGLDR-----LMNPLI--DOVLYLNT 312
 DB 423 IHLRTGDERSTGIYHFDTKPL--KLTHSWQTRNSIGLPPKLTLEPTTGDDHPTLPAA 480
 QY 313 QNOSGSAQNKDLLFSRGSAPGMSVOPKMLPGPCYQGRVSKTKTDNNNSNFTWTGASKY 372
 DB 481 NTRKGVHQITNNSYTEAT--AIRP-----AQGVNTPMNFYSNGGPF 522
 QY 373 NLNGRESIINP-GTMAHKKDEDEKFPMSGVMIPKESAGASNTALD--NMATIDEE 427
 DB 523 -----LTPIVPTADTQYNDDEPN-----GAIRPTMDOYOHGLTTSQ 559
 QY 428 EIK--ATNP-----VATSRFGTVA-VNFOSS-----TDPATG--DVHAMGALP----- 466
 DB 560 ELERYTFNQSKGRAPKQOFNOQAPLNLNTNNGTLTSPDPIGGSNNHFMNTLWTYGP 619
 QY 467 -GMWQDRDLYLGPIWAKIPTH-GHHPSPLMGFGFLK-NPPQILIKNTVPANP 522
 DB 620 LFLANNTAVFPNGQIMDKELDLPRLH--VTPAFCKNNPQGLFVKIAP--NLT 673
 QY 523 AEFSA-TRFASFTQSTGOVSVEIEMELQENSKRMNEVQYTSYASANDPFDVNN 581
 DB 674 DDFNADSPQOPRIITYSNFMWKGTLTLFTAKMSSNMNNDIQOHTT-----TAENI 723

QY 582 GLYTERPPIG 591
DB 724 GKTI-PTNIG 732

RESULT 5
COAL_PAVP9 STANDARD; PRT; 729 AA.

AC P3484;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Porcine parvovirus (strain 90HS) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=33725;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69319168; PubMed=2750278;
RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;
RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";
RL Virus Res. 13:79-86(1989).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
DR PIR; A60006; A60006.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KM Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT 151 729
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 729 AA; 80938 MM; 92538BFP9A0C78B6 CRC64;

Query Match 7.8%; Score 254.5; DB 1; Length 729;
Best Local Similarity 22.7%; Pred. No. 1.6e-09;
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

QY 9 VEGSPQRPDSSSGIGK-----TGQOPAKRLNFGQDSESVDPQPLGEPATP 58
DB 109 VRBRPRHPSKPRGKRPAPRHITINLAKKAKTSMTNSMSENENPINAAGTEL- 167
QY 59 AAVGPTTMASSGGAPMADNNEGADGVGNASNMH--CDSTWLGD--RVITTSRTWALP 113
DB 168 SATNENSGGGGGG-----GGRGAGVGVSSTGFSNNQTEFYLGGLVRIATHASRLHLN 223
QY 114 TYNHNLKQIS--SASTGAS--NDWHYFGYSTPMGCFDNRHCHESPRDQRLNN 167
DB 224 MPEHETKRIHLVNLNSESAGVAGQWODAHHTQVTPWMLIDKAWGVNFPADWOLISNM 283
QY 168 WGFPRKLNFLNFIQVKEVT--TNDGVTTIANLITSTVOVFSDESEYQLPYVLSAHQG 224
DB 284 TEINLVSEFQEIIFNVLKITTESATSPPTKYNNDDLASLWALDITNTLPTPTAAARSE 343
QY 225 CLRPFPADVEMIPQYGL-----TLNNGSQAVGRS-----SFYCLE-YFPS 264
DB 344 TLGFPYPLPTKPTQYRYLSTRNLNPTLYTGSGQQTDSIGTGLHSDIMEYTIENAVPI 403
QY 265 QMLRTGNPF--TFSTFEFVPHSSYASQSILDR-----LAMPPL--DOYLYLNRQ 313
DB 404 HLNTGDEFSFGIYHFTKPL--KLTHSWQTNRLGLPKLLTPTTEGQHPEGLPAAN 461

QY 314 NQSGAQNKDILFRSGSPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSFTWTGASKYN 373
DB 462 TRKGHYGTMMNSYTEAT-----AIRP-----AQGVYTPNPFNSNGPF- 502
QY 374 LNGRESTINP--GTAMASHKDEDEKFFPMGSMVIFPKESAGASNTALDNVMTDESEIRA 431
DB 503 -----LTFIVPTAOTVYVDE-----PNCALFTMGYOHQULTS-----SOELERYT 545
QY 432 TNP-----VATERFGTVV-VNFQSS-----TDPATG--DVHAGALP-----GVMWQ 471
DB 546 FPNQSKGRAPKQAFNQAFNLNNTNNTGLPSPDIPGKRPVNHFNMTLNTYGPPLALNN 605
QY 472 DRDVLQGPITWAKIPTH--GHPSPPLMGFGK-NPPQILIKTTPVAPNPAEFA- 527
DB 606 TAVFPNGQWDELDLTKPRH---VTAFFVCANPPQQLPVKIAF---NLTDENAD 659
QY 528 TKFASFTQYSTQGVSEIEIEMELQKNSKRNPEVQTSNYAKSANVDFVNDNGLYTER 587
DB 660 SPOQPRITVSNFMWKGTLFTAKMRSNNMWNPIQOHTT-----TAENIGNYI-P 708
QY 588 RPIG 591
DB 709 TNIG 712

RESULT 6

COAL_PAVPN STANDARD; PRT; 729 AA.

AC P18346; Q89816;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2] (Version 1).
OS Porcine parvovirus (strain NADU-2) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10797;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021005; PubMed=2219713;
RA Vaudevacharya J., Basak S., Srinivas R.V., Compans R.W.;
RT "The complete nucleotide sequence of an infectious clone of porcine
parvovirus, strain NADU-2.";
RL Virology 178:611-616(1990).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=90085785; PubMed=2596019;
RA Vaudevacharya J., Basak S., Srinivas R.V., Compans R.W.;
RT "Nucleotide sequence analysis of the capsid genes and the right-hand
terminal palindromic of porcine parvovirus, strain NADU-2.";
RL Virology 173:368-377(1989).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M38367; AAA46919.1;
CC EMBL; M38367; AAA46921.1;
CC EMBL; M32787; AAA46917.1;
CC EMBL; M32787; AAA46918.1;
CC PIR; B33743; VCPVNA.
CC HSSP; P30129; 4DPV.
CC InterPro; IPR001403; Parvo_coat.
CC Pfam; PF00740; Parvo_coat; 1.
KM Coat protein; Glycoprotein.


```

FT CHAIN 1 729 COAT PROTEIN VP1.
FT CHAIN 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 175 190 GLY-RICH.
SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80DB42 CRC64;

Query Match 7.8%; Score 254.5; DB 1; Length 729;
Best Local Similarity 22.7%; Pred. No. 1.6e-09;
Matches 152; Conservative 94; Mismatches 272; Indels 151; Gaps 34;

QY 9 VEGSPPEPSSSGIGK-----TGQPAKKRLNFGQDSESVDPDQGEPPATP 58
DB 109 VRSPRKHPSKPPGKRPAPRHIFINLAKKAKGTSTNSNSSEVQHNPIAGTEL- 167
QY 59 AAVGPTTMAAGGAPADNNEGADGVGNASG--NMHCDSTWLDG--RVITSTRTPALP 113
DB 168 SATGNSSGGGGGGG---GGRGAGVGVSTGTENNQTEFOYLGELVRITAHASRLIHIN 223
QY 114 TYNHLLYKOIS--SASTGAS---NDNHFGYSTPMGYDPDNFCHESPRDQRLINN 167
DB 224 MPEHETRYKRIHVLNSESAGVAGQWODAHQWTPMSLIDANAMGWENFADQQLISNM 283
QY 168 WGRFRKRLNFKLNIQVKEVT---TNDGVTTIANNTSTVQVSDSEYQLPYVLGSAHQG 224
DB 284 TEINLVSEGEIENNVVLKTTESATSPPTKIYNNDLTASLMAVLDNTNLTPIYPAARSE 343
QY 225 CLPPPADVEMIPQYGYL-----TLNNSQAVGRS-----SFYCLE-YFPS 264
DB 344 TLGTFYPLPTKPYQRYYLSCIRNLNAPPTYGSSQOITDSIQGLSHDIMEFTIENAVPI 403
QY 265 QMLRTGNPF-TFSYTFEEVPHSSVYASQSLDR-----LNPPLI--DOYLYLYNRQ 313
DB 404 HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNBSLGLPRLKLEPTTEGQHPTGLPAAN 461
QY 314 NQSGAONKLLFSRSGPAGMVLPCPCYRQORVSKTKTDNNNSNFTWTGASKYN 373
DB 462 TRKGYNQTNINSYTEAR---AIRP-----AQVGNTPYNNFEVSNNGPFF- 502
QY 374 LNGRESIINP--GTAMASHKDEDEKFFPMGSMVIFGESAGASNTALD---NVMITDEE 428
DB 503 -----LTPIVPTADQYNDDEPN-----GAIKFTWDYOHGHLTSSQE 540
QY 429 IK--ATNP-----VATERFGTVA--VNFQSSS-----TDPATG--DVHAMGALP----- 466
DB 541 LERVTYFPOSGCGAPRQFNOQAPLLENTNNTLLPSDPIGKSMHMMNTLNTYGPL 600
QY 467 GATWQDRDVLQSPITAKIPHTD--GHFHSPLMGSGFLK--NPPQILLIKTTPANRPA 523
DB 601 TALNNTAPVFPNGQIMKELDIDKPRILH---VTAIPVCKANNPPGQIFVIAIP---NLTD 654
QY 524 EFGA-TKFAFTQYSTGVSVLEIWELOKENSKRANPEVOYTSNAXKANVDFYDANG 562
DB 655 DFNADSPQQRILITYSNFMWKGLTFTAKRSMNMPDIOHHT-----TANIG 704
QY 583 LYTEPRIG 591
DB 705 NYI-PTNIG 712

RESULT 7
ID COAL_PAVPK STANDARD; PRT; 729 AA.
AC P52501;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coate protein VP1 (Contains: Coat protein VP2).
OS Porcine parvovirus (strain Kresse) (PPV).
OC Viruses; ssRNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
ON NCBI_TaxID=73487;
RX MEDLINE=96183900; PubMed=8642680;
RA Bergeron J, Hebert B, Tijssen P;
RT "Genome organization of the Kresse strain of porcine parvovirus: identification of the allosteric determinant and comparison with those of MAD-2 and field isolates."
RT J. Virol. 70:2508-2515(1996).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL: U44978; AAC0230.1; -.
CC EMBL: U44978; AAC0231.1; -.
CC HSSP: P30129; 4DPV.
CC InterPro: IPR001403; Parvo coat.
CC Pfam: PF00740; Parvo coat: 1.
CC KW Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT CHAIN 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 729 AA; 80835 MW; B6345BFA058BA1F6 CRC64;

Query Match 7.8%; Score 252.5; DB 1; Length 729;
Best Local Similarity 22.7%; Pred. No. 2.2e-09;
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

QY 9 VEGSPPEPSSSGIGK-----TGQPAKKRLNFGQDSESVDPDQGEPPATP 58
DB 109 VRSPRKHPSKPPGKRPAPRHIFINLAKKAKGTSTNSNSSEVQHNPIAGTEL- 167
QY 59 AAVGPTTMAAGGAPADNNEGADGVGNASGMH--CDSTWLDG--RVITSTRTPALP 113
DB 168 SATGNSSGGGGGGG---GGRGAGVGVSTGTENNQTEFOYLGELVRITAHASRLIHIN 223
QY 114 TYNHLLYKOIS--SASTGAS---NDNHFGYSTPMGYDPDNFCHESPRDQRLINN 167
DB 224 MPEHETRYKRIHVLNSESAGVAGQWODAHQWTPMSLIDANAMGWENFADQQLISNM 283
QY 168 WGRFRKRLNFKLNIQVKEVT---TNDGVTTIANNTSTVQVSDSEYQLPYVLGSAHQG 224
DB 284 TEINLVSEGEIENNVVLKTTESATSPPTKIYNNDLTASLMAVLDNTNLTPIYPAARSE 343
QY 225 CLPPPADVEMIPQYGYL-----TLNNSQAVGRS-----SFYCLE-YFPS 264
DB 344 TLGTFYPLPTKPYQRYYLSCIRNLNAPPTYGSSQOITDSIQGLSHDIMEFTIENAVPI 403
QY 265 QMLRTGNPF-TFSYTFEEVPHSSVYASQSLDR-----LNPPLI--DOYLYLYNRQ 313
DB 404 HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNBSLGLPRLKLEPTTEGQHPTGLPAAN 461

```

```

QY 314 NQSSAQNKDLFERGSPAGMSVOPKMWLPFCYRQQRVSKTITDNNNSFTWGAQKYN 373
DB 462 TRKYHGTJNNSTYAT- -AIRP- - - - - - - - - - - - - - - - - - - - -
QY 374 LMGSESTINP- -GTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALDNWITDEEIK 431
DB 503 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 432 TNP- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
DB 546 FNPQSKGCRAPKQCFNQAAPLNTLNTNNTLLPSPDPIGKKNHGFNNTLNTYGPDLTALNN 605
QY 472 DROVYLOGPIWAKLPHTD- -GHFHPSPLMGSGRGLK- -NPPQILIKNTVPANPAEBSA- 527
DB 606 TAPFENQGLWDEKLDLTKRPLH- - - - - - - - - - - - - - - - - - - - -
QY 528 TKFASFTQYSTGVSEIEMELQKNSKRNPEVQYTSNYAKSANYDFVDDNNGLYTER 587
DB 660 SPQQRPIITYSNFMWKGTLFTAGRBSNMNPIQOHTT- - - - - - - - - - - - - - -
QY 588 RPIG 591
DB 709 TNIG 712

RESULT 8
COAT_PAVH STANDARD; PRT; 722 AA.
ID PO3136;
AC P03136;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2].
OS Hamster parvovirus H1.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10799;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85112183; PubMed=6823009;
RA Rhode S.L. III, Paradiso P.R.;
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its
RT genes by hybrid-arrested translation.";
RL J. Virol. 45:173-184 (1983).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X01457; CAB57285.1; ALT_SEQ.
DR PIR; A03699; VCPVW2.
DR HSSP; P07302; IMVM.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KM Coat protein; Glycoprotein.
FT CHAIN 1 722 COAT PROTEIN VP1.
FT CARBOHYD 131 722 COAT PROTEIN VP2.
FT CARBOHYD 178 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 647 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 675 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 155 170 GLY-RICH.
SQ SEQUENCE 722 AA; 79737 MW; 6BB678391AASDC31 CRC64;
Query Match 7.5%; Score 244; DB 1; Length 722;
Best Local Similarity 21.4%; Pred. No. 8e-09;
Matches 142; Conservative 94; Mismatches 267; Indels 160; Gaps 29;

```

```

QY 8 PVESQPEPDSSSGIGKTGO- - - - - - - - - - - - - - - - - - - - -
DB 82 PKLSTDSER- -GTGVSPPGKRTKPPAHIFYNQAPAKKRSALAAOQRTILMSGTETNP 140
QY 55 PATPAVGPPTTMSGGGAPADNNEGADVGNAGMWCDSW- - - - - - - - - - - - - - -
DB 141 DTGIANRVERSRADGGGS- -SGGGSGGGGIGVSTGYNDQNTYKPLDGGWVEITAAHSRL 199
QY 113 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
DB 200 LHLGMPSENYCYVTHANNQTTGHGTVKGNMAVYDTHQITM- -TPWSLYDANAGWVFQPS 258
QY 159 DMORLIINNMGRPKLNFLENIQYKEVTT- - - - - - - - - - - - - - - - - - - - -
DB 259 DMQFIQMSMSLMDLSQELFNVVYKTYEQAGADAIKYNNNDITACMTALDSNNT 318
QY 214 LPVLAGSAHQGL- - - - - - - - - - - - - - - - - - - - - - - - - - -
DB 319 LPYTPAAQTSSETLGFYPMKPTAPARYRYFFMRQLSVTSNSAEGTQITDTIGEPQALN 378
QY 254 SSFYCLE- -YPPSQMLRTGNFT- -STFEVFPHSYA- - - - - - - - - - - - - - -
DB 379 SOFTIETLPIPLRTGDEFTTGYIFNTDPLKLTHTWQTNHLLACLOGITDLPSTDA 438
QY 306 LYYLNRTONQSGAQNKDLF- - - - - - - - - - - - - - - - - - - - -
DB 439 TASILANGDRFGSTQYQNVVYTEALRTPAQIGFMQPHDNFANNGPFRVYVP- - - - -
QY 343 PGBCYRQQRVSKTITDNNNSFTWGAQKYNLGR- - - - - - - - - - - - - - -
DB 495 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 391 KDEDEKFFPMGSMVIFGKESAGASNTALDNWITDEEIK- - - - - - - - - - -
DB 542 RDTARCV- - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 450 SSTDPATGVHANGALPGMWQODRVYLOGPIWAK- - - - - - - - - - -
DB 592 PHEDP- - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 507 PPQILIKNTVPANPAEF- -SATKFSFTQYSTGVSEIEMELQKNSKRNPEVQY 564
DB 629 PGOLFVHLGP- - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 565 TSN 567
DB 685 TTD 687

RESULT 9
COAT_PAVL3 STANDARD; PRT; 587 AA.
ID PO36310;
AC P36310;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2].
OS Parvovirus Lu111.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=35339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83297126; PubMed=8517025;
RA Diffoot N., Chen K.C., Bates R.C., Lederma M.;
RT "The complete nucleotide sequence of parvovirus Lu111 and
RT localization of a unique sequence possibly responsible for its
RT encapsidation pattern.";
RL Virology 192:339-345 (1993).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


[3]
 RN X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.
 RA J. Llamas-Saiz A.L., Agbandje-Mckenna M., Winkler W.R., Bratton J.,
 RA Tattersall P., Rosemann M.G.,
 RT "Structure determination of Minute Virus of mice."
 RL Acta Crystallogr. D 53:93-100(1997).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X02481; CAB46507.1; -;
 DR EMBL; X02481; CAB46508.1; -;
 DR EMBL; M12032; AAA69569.1; ALT_INIT.
 DR PIR; B23008; VCPVIM.
 DR PDB; 1MWI; 25-FEB-98.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein; 3D-structure.
 FT CHAIN 1 718 COAT PROTEIN VP1.
 FT 1 718 COAT PROTEIN VP2.
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 502 502 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 157 172 GLY-RICH.
 FT CONFLICT 144 144 A -> G (IN REF. 2).
 SQ SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;
 Query Match 6.6%; Score 214.5; DB 1; Length 718;
 Best Local Similarity 21.0%; Pred. No. 7.1e-07;
 Matches 142; Conservative 89; Mismatches 273; Indels 171; Gaps 30;
 QY 8 PVESGPOEPDSSSGIGTKG-----QDPKRLNFG--QTGSESVDDPQPLG 52
 DB PKLATSEP-GTSGVSRAGKRTTRPAYIFINQARAKKLTSSAAQSSQTMSTGTSQPDG 141
 QY 53 -----EPPTPAVGGPTTMAAGGAPMANNNEGADGVNAGNWHDS--TWLGD--R 101
 DB 142 GNAVSAARVERADGPG--GSGGG-----GSGGGGVSTGSDYNDQTYRRLGDEGWE 193
 QY 102 VITTSRTTALPTYNHLYKQI-----SSASTGASNDHRYGYSTPMGVDFNRFCH 154
 DB 194 ITLALTRVLNLPKSNKGRIRVHNTDTISYKGNNAKDAHQIWTPLSLVDANAMGW 253
 QY 155 FSPRDWQRLINNNGFRKRLNFKLNIQYKEVTTND---GVTTANLSTFVQFSDS 210
 DB 254 LQSDWQYICNTMSQLNLVSLDOEIFVVLKTVEDSGGQAIKYNNDLTACMVAVDS 313
 QY 211 EYGLPYVLSAGHGGCLPPPADVFMIOGY-----LTLNN-----GQAV 251
 DB 314 NNILPTPAANSMTETGFPWKPTIASPYRYFVCDRLDSVTYENOGTIEHNVMGPKG 373
 QY 252 GRSSFCLEYFSP--OMLRTGNF--TFSYTEEVEVPHSSVHOSLDILNM--PLIDPYLY 307
 DB 374 MNSQFTIENTQITLIRTDDEFATGYTFDTPNV--KLHTMOTNRQLQOPPLSTF-- 429
 QY 308 YLNRTONQSSAONKDLIFSRGSPAGMSVOPKWL-----P 343
 DB 430 --PEADTDAGT-----LTQSGRHGATOMEVNVVSAIRTPRAYGFCOPHNDFEASRA 481
 QY 344 GPCYRQORVSKTKTDNNNSFTWGASKYMLNGRESIINPTAMASHKDDKDR----- 397
 DB 482 GP-FAAPKIVADVTQGVDRKANGSVYSYKQGHENNAHAPAPERTYTDETNFGSGRDT 540
 QY 398 ---FPMGVMIFGKSGAGASNTALDNVMTDEEIKATNPATERFQTAVNFGS--SST 452
 DB 159 TGSGNGSGGG-----GGSGGVGISTGTFFNNQTEFFKFLNMGV--ETANSRLVHLN 210

DB 541 RDGFQASAPLVPPPIINGLITNA-----NFIGTKN-----DIHFSNVENS 581
 QY 453 DPATGVHAMGALPGWMDRVDYLOOPTMAK---IPHTDGHHPRLMGFGLKNPPQ 509
 DB 582 GPITAFSH-----PSPYPOGQIWDKLELDHPRRLHITAPFV---CKNNAPOG 627
 QY 510 ILTKNTP--VPANPAEFATKFSFITYOSTGVSEVEIEMELQKNSKRWNPVOYTSN 567
 DB 628 MLVRLQPNLTDQYDPPNGATISRLVYTGTFPMKGLMRAKLA---NTWNVYQ---- 679
 QY 568 YAKSANVDFTVNNG 582
 DB 680 -----VSVEDNG 686
 RESULT 13
 COAT MEVA STANDARD; PRT; 722 AA.
 ID P27437;
 AC 01-AUG-1992 (Rel. 23, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2].
 OS Mink enteritis virus (strain Abashiri) (MEV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=10793;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91202123; PubMed=2016597;
 RA Karitsumari T., Horituchi M., Hama E., Yaguchi K., Ishiguro N.,
 RA Goto H., Shinagawa M.;
 RT "Construction and nucleotide sequence analysis of an infectious DNA
 RT clone of the autonomous parvovirus, mink enteritis virus."
 RL J. Gen. Virol. 72:867-875(1991).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00765; BAA00663.1; -;
 DR PIR; B38350; VCPVME.
 DR HSSP; P30129; 4DPV.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 722 COAT PROTEIN VP1.
 FT 1 722 COAT PROTEIN VP2.
 FT DOMAIN 139 722 GLY-RICH.
 FT CARBOHYD 160 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 722 AA; 79823 MW; 9DADBACAB2EF9F622 CRC64;
 Query Match 6.6%; Score 214; DB 1; Length 722;
 Best Local Similarity 20.2%; Pred. No. 7.8e-07;
 Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;
 QY 7 RPVEGSPQEPDSSSGIGTKGQCPAKKRLNFGQGDSESV-----DQPLGEPPA--TPAA 60
 DB 105 KPTKRSKPPPHITINLAK-----KKAGAGQYKRNLAHMSGAVQPDGQGAIVANERA 158
 QY 61 VGPPTMASGGGAPMADNNEGADGVNAGNWH-----CDSTWLDGRVITTSRTWALP 113
 DB 159 TGSGNGSGGG-----GGSGGVGISTGTFFNNQTEFFKFLNMGV--ETANSRLVHLN 210

```

QY 114 TYNHLYKQI-----SSASTGASNDNHFGSTPMGYPDFRPHCHSPRDMQRLINN 166
DB 211 MPBSENKRVVANNMDKTAVKGNMALLDTHVOIVTPMSLVDAAMGVWPFNGMQLVNT 270
QY 167 NMGRPRKLNFKLFNIQVKEVT---TNDGVTTIANNLSTVQVPSDEYOLPYVLGSAHQ 223
DB 271 MSELHLVSEBOEINNVVLKTVSESATOPRTKVYNNDLTASLMLVALDSNNTMPTFPALMRS 330
QY 224 GCL-----PPFPA-----DVPMIPQYGLTLNNGSAVARS-----SFYCL 259
DB 331 ETLAGFYPMKPTIPPMRYYPFQMDTLLPSH-----TGSGTPTNIGHGDPDDVQFYTI 384
QY 260 E-YRPSQMLRTGNF--TFSTYFESEVPHSSVAHSQSILRLMN--PLIDQVLYLNRQNG 315
DB 385 ENSVPHLRTGDEFATGTFPFDCKP--CRLTHWQTRALGLRP-----FLNSLPQS 435
QY 316 SGSANOKDLFSGSPAGMS-----VQPKN-MLP3PCYRQQRVSK----- 354
DB 436 EGATNPGDIGVQDKRGVLTQMGNTDYITETATIRPAVEGSAVYSEASTGQPFKPTI 495
QY 355 -----TKTDNNNSNFTWTGASKY---NLNGRESIINPGT---AMASHKODEDKFPPMS 401
DB 496 AAGRGAGQTDENQAA---DGDPRYAFGRHOGKTTTGETPERFTYIAHODT----- 544
QY 402 GWMIFGKESAGASNTALD--NVMITDEBEIKATNPVATERFGTVAVNFQSSSTDPAQDVH 460
DB 545 -----GRYPBGDWIQININFLPVTNDVLLPTDPIG---GKTGINV--TNINFTYGPILT 593
QY 461 AMGALPGMWQODRDVYLQGPIMAKIPHTDGHFHPSPLMGGFGLK-----NPP 507
DB 594 ALNNVP-----PVYPNGQIMDKPEPTD-----LKPRLHVNAPFVCONNCP 633
QY 508 POLIKNTKTPVAN--PPAESATKFAFITQYSTGVSVIEIWELOKENSKEKNPEVOY 564
DB 634 GQLPVKAPLITNEDPDASANMSR---IVTYSDFWKKGLVFKALKLRASHTNPIQOM 689
QY 565 TSNYAKSANVDFTVNNGLYTEPRPIG 591
DB 690 SIN-----VDNQPNYL-PNNIG 705

RESULT 14
COAT_PPV19
ID_COAT_PPV19 STANDARD; PRT; 727 AA.
AC 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Feline panleukopenia virus (strain 193) (FPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10787;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=193/70;
RX MEDLINE=91073139; PubMed=2174965;
RA Martyn J.C., Davidson B.E., Studdert M.J.;
RT "Nucleotide sequence of feline panleukopenia virus: comparison with
RT canine parvovirus identifies host-specific differences.";
RL J. Gen. Virol. 71:2747-2753(1990).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=CU-4;
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine
RT parvovirus and feline panleukopenia virus using infectious plasmid
RT clones.";
RL Virology 183:195-205(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC
DR EMBL: X55115; CA38911.1; -
DR EMBL: M38246; AAC37928.1; -
DR EMBL: M38246; AAC37929.1; -
DR F1R; B3608; VCPVP.
DR HSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 727 COAT PROTEIN VP1.
FT DOMAIN 144 727 COAT PROTEIN VP2.
FT FT 165 190 GLY-RICH.
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 165 180 GLY-RICH.
SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621F5S CRC64;

Query Match 6.4%; Score 209; DB 1; Length 727;
Best Local Similarity 20.1%; Pred. No. 1,7e-06;
Matches 138; Conservative 101; Mismatches 260; Indels 188; Gaps 34;

QY 7 RPYVSGQPEPSSSGIGKGTQOPAKKRLNFGOTGDSSEVP-----DPOPLEPPA--TPAA 60
DB 110 KPTKRSKPPHITINLAK-----KKKAGQYKRDNLAMSGANVPDGGQAVNRERA 163
QY 61 VGPPTMASGGGAPWADNNEGADVGNASGNW-----CDSTWLGDRVITTSRTWALP 113
DB 164 TSGSNGSGGG-----CGSGGAGVIGSTGFNNQTEKFLENGWV--EITANSRLVHLN 215
QY 114 TYNHLYKQI-----SSASTGASNDNHFGSTPMGYPDFRPHCHSPRDMQRLINN 166
DB 216 MPBSENKRVVANNMDKTAVKGNMALLDTHVOIVTPMSLVDAAMGVWPFNGMQLVNT 275
QY 167 NMGRPRKLNFKLFNIQVKEVT---TNDGVTTIANNLSTVQVPSDEYOLPYVLGSAHQ 223
DB 276 MSELHLVSEBOEINNVVLKTVSESATOPRTKVYNNDLTASLMLVALDSNNTMPTFPALMRS 335
QY 224 GCL-----PPFPA-----DVPMIPQYGLTLNNGSAVARS-----SFYCL 259
DB 336 ETLAGFYPMKPTIPPMRYYPFQMDRLLPSH-----TGSGTPTNIGHGDPDDVQFYTI 389
QY 260 E-YRPSQMLRTGNF--TFSTYFESEVPHSSVAHSQSILRLMN--PLIDQVLYLNRQNG 315
DB 390 ENSVPHLRTGDEFATGTFPFDCKP--CRLTHWQTRALGLRP-----FLNSLPQS 440
QY 316 SGSANOKDLFSGSPAGMS-----VQPKN-MLP3PCYRQQRVSK----- 354
DB 441 EGATNPGDIGVQDKRGVLTQMGNTDYITETATIRPAVEGSAVYSEASTGQPFKPTI 500
QY 355 -----TKTDNNNSNFTWTGASKY---NLNGRESIINPGT---AMASHKODEDKFPPMS 401
DB 501 AAGRGAGQTDENQAA---DGDPRYAFGRHOGKTTTGETPERFTYIAHODT----- 549
QY 402 GWMIFGKESAGASNTALD--NVMITDEBEIKATNPVATERFGTVAVNFQSSSTDPAQDVH 460
DB 550 -----GRYPBGDWIQININFLPVTNDVLLPTDPIG---GKTGINV--TNINFTYGPILT 598
QY 461 AMGALPGMWQODRDVYLQGPIMAKIPHTDGHFHPSPLMGGFGLK-----NPP 507
DB 599 ALNNVP-----PVYPNGQIMDKPEPTD-----LKPRLHVNAPFVCONNCP 638
QY 508 POLIKNTKTPVAN--PPAESATKFAFITQYSTGVSVIEIWELOKENSKEKNPEVOY 564

```

Db 639 GOLFVKAAPULITEYDDASANNMR-----IVTYSDFWMMKGLVFKAKLRASHWTNPIQOM 634
 QY 565 TSNVAKSANDVFTYDNNGLTEPRPIG 591
 Db 695 SIN-----VDNQFNIV-PNNIG 710

RESULT 15

COAT_PAVCB
 ID COAT_PAVCB STANDARD; PRT; 727 AA.

AC 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [contains: Coat protein VP2].
 OS Canine parvovirus (strain B) (CPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI TaxID=59284;

RP SEQUENCE FROM N.A.
 RX MEDLINE=91272479; PubMed=1647068;

RA Barish C.R.;
 RT Mapping specific functions in the capsid structure of canine parvovirus and feline panleukopenia virus using infectious plasmid clones.";

RL Virology 183:195-205(1991).

CC -I- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, OF A COMBINATION OF VP2, VP3, AND SOME VP1.

CC -I- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M38245; AAB02799.1; -;
 DR EMBL; M38245; AAB02800.1; -;

DR HSSP; P310129; 4DPV.

DR InterPro; IPR001403; Parvo coat.

DR Pfam; PF00740; Parvo coat; 1.

KM Coat protein; Glycoprotein.

FT CHAIN 1 727 COAT PROTEIN VP1.

FT CARBOHYD 144 727 COAT PROTEIN VP2.

FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.4%; Score 208; DB 1; Length 727;
 Best Local Similarity 19.8%; Pred. No. 2e-06;
 Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;

QY 7 RPEVQSPDEPSSSGIGKTCQAPKRLNFGQTSSEVP---DPOPLGEPPA--TPAA 60
 Db 110 KPTKRSKPPHIFINLAK-----KKKAGAGVYRDLALPMSDGAVOPDGGCPAVNNERA 163
 QY 61 VEPPTMASGGAPMADNNEGADVGNASGWH-----CDSTWLGDRVITTSRTWALP 113
 Db 164 TSGSGNSGGGGG-----GGSGGAGISTGTFFNQTEKFFLENGWV--EITANSSRLVHLN 215
 QY 114 TYNHLYKO-----SSASTGASNDNHFGYSTPWGFDFNRFHCHSPRDMORLINN 166
 Db 216 MPSESENTRRVVNNMKTAVNGMALLDDIHAQIVTFPWSLVDANAMGVWENPBGDMQILTNT 275
 QY 167 NMGFRPKRLNFKLNIQVKEVT--TNDGVTTIANNLTSTVQVPSDEYQLPYVLGSAHQ 223

Db 276 MSELHLVSFEQELFNVVLKTVSSSATQPTKYVNNDLTASLWALBNSNMTPTPAAMS 335
 QY 224 GCL-----PPPPA-----DVEMIPQYGITLNGSOAVRS-----SFYCL 259
 Db 336 ETLGFYPMKPTIPTPMRYFFQMDRLIPSH-----TGTSGPTNHYHGTDDPDVQFYTI 389
 QY 260 E-YFPSQMLRTGNF--TFSTYFEEVPHSSVYASQSLDLMN--PLIDQYLYVINTQNG 315
 Db 390 ENSVPVHLRTGDEFAITGTFPDKP--CRLTHTWQTNALGLP-----FLNSLPQS 440
 QY 316 SGSAQNKDLFFSGSPAGMSVQPKMLPGPCYKQRFVSRTKTDNNNSN-----TW 366
 Db 441 EGATNFGDI-----GV-----QDQKRGVQTQMGNTVITTEATIKRPAE 478
 QY 367 TGASKYNLNGRESITNP-----GTAMASHKDEDEKFFPMGCVIIFKESAGASNTALD-- 419
 Db 479 VGYSAPYSPFASSTQGPFKTPIAAGRGAGQTDENQADGNPRYAFGRHGOAKTTTIGETP 538
 QY 420 -----NVMITDEEIKATNPVATERGTVAVNFQSSSTDP 454
 Db 539 ERETTYAHQDTGRYPRGDMITQINENFLPTNDNVLLPTDPIC---GKTGINY--TNIN 592
 QY 455 ATGDVHAMGALPGMWQDRDYYLQGPIMAKIPTDGHFHPSPIMGFGLK-----504
 Db 593 TYGFLTALNNVP-----PYVNGQIMDKFPTD-----LKPRLHVNAPEV 632
 QY 505 ---NPPQILLIKTTPYPA--PPAEFSATKRASTFTQYSTGQVSVIEIMELCKENSKW 558
 Db 633 CONNCGQLFVKVAPULITEYDDASANNMR-----IVTYSDFWMMKGLVFKAKLRASHTW 688
 QY 559 NPEVQTSNVAKSANDVFTYDNNGLTEPRPIG 591
 Db 689 NPIQMSIN-----VDNQFNIV-PSNIG 710

Search completed: July 17, 2003, 18:32:05
 Job time : 14.8582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 Search time 44.5484 Seconds
(without alignments)
2770.518 Million cell updates/sec

Title: US-09-807-802a-15

Perfect score: 3251

Sequence: 1 TABGKRPEVQSPQEPDSSS.....NNGLYTEPRPIGRVILTRPL 599

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251	100.0	736	12	Q9WBP8
2	3229	99.3	736	12	Q56137
3	2832	87.1	736	12	Q56139
4	2815	86.6	736	12	Q65311
5	2764.5	85.0	735	12	Q56552
6	2759.5	84.9	598	12	Q56553
7	2481.5	76.3	533	12	Q92917
8	1830.5	56.3	734	12	Q41855
9	1717	52.7	732	12	Q67666
10	1714	52.7	732	12	Q83290
11	1712	52.7	587	12	Q67667
12	1712	52.7	732	12	Q8V395
13	1702	52.4	732	12	Q65444
14	1701	52.3	732	12	Q83289
15	1700.5	52.3	724	12	Q9Y1J1
16	1698	52.2	534	12	Q67668

17	1695	52.1	587	12	Q65445	Q65445	barbarie du
18	1676	51.6	534	12	Q65446	Q65446	barbarie du
19	1540	47.4	676	12	Q67672	Q67672	goose parvo
20	715	22.0	179	12	Q9WA24	Q9WA24	goose parvo
21	715	22.0	179	12	Q9W8U7	Q9W8U7	duck parvo
22	711	21.9	179	12	Q9WN18	Q9WN18	duck parvo
23	709	21.8	179	12	Q9WN19	Q9WN19	duck parvo
24	707	21.7	179	12	Q9WN20	Q9WN20	goose parvo
25	632.5	19.5	703	12	Q8QCV5	Q8QCV5	minute viru
26	618	19.0	571	12	Q8QCV4	Q8QCV4	minute viru
27	545.5	16.8	947	12	Q918U9	Q918U9	bovine parv
28	508	15.6	781	12	Q9PZT0	Q9PZT0	human parvo
29	501	15.4	554	12	Q9PZS9	Q9PZS9	human parvo
30	501	15.4	781	12	Q9JGP8	Q9JGP8	human parvo
31	500.5	15.4	773	12	Q913X1	Q913X1	human parvo
32	499.5	15.4	785	12	Q910X4	Q910X4	pig-tailed
33	499	15.3	781	12	P89318	P89318	human parvo
34	499	15.3	781	12	P89319	P89319	human parvo
35	498.5	15.3	781	12	P89317	P89317	human parvo
36	497.5	15.3	781	12	Q912B8	Q912B8	human eryth
37	497	15.3	769	12	Q9PZT4	Q9PZT4	human parvo
38	497	15.3	781	12	Q9JGSO	Q9JGSO	human parvo
39	497	15.3	781	12	P90221	P90221	human parvo
40	497	15.3	781	12	P90222	P90222	human parvo
41	497	15.3	781	12	P90223	P90223	human parvo
42	497	15.3	781	12	P90224	P90224	human parvo
43	497	15.3	781	12	Q85191	Q85191	human parvo
44	497	15.3	781	12	Q85117	Q85117	human parvo
45	497	15.3	781	12	Q90200	Q90200	human parvo

ALIGNMENTS

RESULT 1
ID Q9WBP8 PRELIMINARY; PRT; 736 AA.
AC Q9WBP8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein.
OS adeno-associated virus 1.
OC Viruses; ssDNA viruses; Parvoviridae; Dependovirus.
OX NCBI_TaxID=85106;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9214338; PubMed=10196295;
RX Xiao W., Chirmule N., Beta S.C., McCullough B., Gao G., Wilson J.M.;
RT "gene therapy vectors based on adeno-associated virus type 1.";
RL J. Virol. 73:3994-4003 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Xiao W., Wilson J.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF063497; AAD27757.1; -;
DR InterPro: IPR001403; Parvo_coat.1.
DR Pfam: PF00740; Parvo_coat.1.
SQ SQUENCE 736 AA; 81375 MW; CFABP9BDS0595 CRC64;
Query Match 100.0%; Score 3251; DB 12; Length 736;
Best Local Similarity 100.0%; Pred. No. 9.5e-225;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TABGKRPEVQSPQEPDSSSGIGKTGQPAKRLNGQTGDSHSDPDLGPAPTPAA 60
DB 138 TABGKRPEVQSPQEPDSSSGIGKTGQPAKRLNGQTGDSHSDPDLGPAPTPAA 197
QY 61 VGTITMASGGAGAMANNMGAGDGVGNASGNWHDSTWLDGRTVITSTRTMALPTNNHLY 120
DB 198 VGTITMASGGAGAMANNMGAGDGVGNASGNWHDSTWLDGRTVITSTRTMALPTNNHLY 257
QY 121 KQTSASATGASNDNHFGYSTPWGYPDFNRHCHFPSPDMORLLNNMGFRPKLFLKLF 180

```
Db 258 KOISSASTGASNDNHFGYSTPMWGFDFNRFCHFSRPMQRLINNMMGFRPKLNFKLF 317
Qy 181 NIOVKEVTTNDGVTIANNTSTVOVPSDEYOLPYVLSAHQGLPPFADVFMIPOYG 240
Db 318 NIOVKEVTTNDGVTIANNTSTVOVPSDEYOLPYVLSAHQGLPPFADVFMIPOYG 377
Qy 241 YLTNNQSAVGRSSFCLEYFSPQMLRTGNNTFSYFEEVPHSSVAHSQSLDRLMNP 300
Db 378 YLTNNQSAVGRSSFCLEYFSPQMLRTGNNTFSYFEEVPHSSVAHSQSLDRLMNP 437
Qy 301 LIDQYLYLNRTONQSSAQNKKLLFSRGSPPAGMSVQPKMLPGPCYRQGRVSKTKTDNN 360
Db 438 LIDQYLYLNRTONQSSAQNKKLLFSRGSPPAGMSVQPKMLPGPCYRQGRVSKTKTDNN 497
Qy 361 NSNFTWTSKATNLRGRESTINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420
Db 498 NSNFTWTSKATNLRGRESTINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 557
Qy 421 VMTDEBEIKATNPVATERGTVAVNFQSSSTDPATGCVHMGALPGMWQDRDYVLOGP 480
Db 558 VMTDEBEIKATNPVATERGTVAVNFQSSSTDPATGCVHMGALPGMWQDRDYVLOGP 617
Qy 481 IWAKIPHTDGHFHPSPLMGFGFLKKNPPQILLKNTVPANPAPFASATKFASTTQYSTG 540
Db 618 IWAKIPHTDGHFHPSPLMGFGFLKKNPPQILLKNTVPANPAPFASATKFASTTQYSTG 677
Qy 541 QVSVEIEMELQKNSKRNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTIRYLTRPL 599
Db 678 QVSVEIEMELQKNSKRNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTIRYLTRPL 736
```

RESULT 2

```
OS6137 PRELIMINARY; PRT; 736 AA.
ID 056137
AC 056137
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
RT (AAV) serotypes other than AAV type 2."
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;
```

Query Match 99.3%; Score 3229; DB 12; Length 736;
Best Local Similarity 99.2%; Pred. No. 3.6e-223;
Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 TAPGKKRPVQSPQEPDSSGIGTKGQOPAKKRLNFGQTDSESVDPQPLGEPAPTPAA 60
Db 138 TAPGKKRPVQSPQEPDSSGIGTKGQOPAKKRLNFGQTDSESVDPQPLGEPAPTPAA 197
Qy 61 VGPTTASGGGAPMADNNEGADGVGNASGNMHCSTWLGDVITTSRTWALPTYNHLY 120
Db 198 VGPTTASGGGAPMADNNEGADGVGNASGNMHCSTWLGDVITTSRTWALPTYNHLY 257
Qy 121 KOISSASTGASNDNHFGYSTPMWGFDFNRFCHFSRPMQRLINNMMGFRPKLNFKLF 180
```

```
Db 258 KOISSASTGASNDNHFGYSTPMWGFDFNRFCHFSRPMQRLINNMMGFRPKLNFKLF 317
Qy 181 NIOVKEVTTNDGVTIANNTSTVOVPSDEYOLPYVLSAHQGLPPFADVFMIPOYG 240
Db 318 NIOVKEVTTNDGVTIANNTSTVOVPSDEYOLPYVLSAHQGLPPFADVFMIPOYG 377
Qy 241 YLTNNQSAVGRSSFCLEYFSPQMLRTGNNTFSYFEEVPHSSVAHSQSLDRLMNP 300
Db 378 YLTNNQSAVGRSSFCLEYFSPQMLRTGNNTFSYFEEVPHSSVAHSQSLDRLMNP 437
Qy 301 LIDQYLYLNRTONQSSAQNKKLLFSRGSPPAGMSVQPKMLPGPCYRQGRVSKTKTDNN 360
Db 438 LIDQYLYLNRTONQSSAQNKKLLFSRGSPPAGMSVQPKMLPGPCYRQGRVSKTKTDNN 497
Qy 361 NSNFTWTSKATNLRGRESTINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 420
Db 498 NSNFTWTSKATNLRGRESTINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDN 557
Qy 421 VMTDEBEIKATNPVATERGTVAVNFQSSSTDPATGCVHMGALPGMWQDRDYVLOGP 480
Db 558 VMTDEBEIKATNPVATERGTVAVNFQSSSTDPATGCVHMGALPGMWQDRDYVLOGP 617
Qy 481 IWAKIPHTDGHFHPSPLMGFGFLKKNPPQILLKNTVPANPAPFASATKFASTTQYSTG 540
Db 618 IWAKIPHTDGHFHPSPLMGFGFLKKNPPQILLKNTVPANPAPFASATKFASTTQYSTG 677
Qy 541 QVSVEIEMELQKNSKRNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTIRYLTRPL 599
Db 678 QVSVEIEMELQKNSKRNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTIRYLTRPL 736
```

RESULT 3

```
OS6139 PRELIMINARY; PRT; 736 AA.
ID 056139
AC 056139
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
RT (AAV) serotypes other than AAV type 2."
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028705; AAB95452.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;
```

Query Match 87.1%; Score 2832; DB 12; Length 736;
Best Local Similarity 85.7%; Pred. No. 1.1e-194;
Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;

```
Qy 1 TAPGKKRPVQSPQEPDSSGIGTKGQOPAKKRLNFGQTDSESVDPQPLGEPAPTPAA 60
Db 138 TAPGKKRPVQSPQEPDSSGIGTKGQOPAKKRLNFGQTDSESVDPQPLGEPAPTPAA 197
Qy 61 VGPTTASGGGAPMADNNEGADGVGNASGNMHCSTWLGDVITTSRTWALPTYNHLY 120
Db 198 VGPTTASGGGAPMADNNEGADGVGNASGNMHCSTWLGDVITTSRTWALPTYNHLY 257
Qy 121 KOISSASTGASNDNHFGYSTPMWGFDFNRFCHFSRPMQRLINNMMGFRPKLNFKLF 180
Db 258 KOISSASTGASNDNHFGYSTPMWGFDFNRFCHFSRPMQRLINNMMGFRPKLNFKLF 316
```

QY 181 NIOKVEYTNQDGVTTIANNLSTVQVFSDEYOLPYVLGSAHQCLPEPPADVEMIPQY 240
DB 317 NIOKVEYTNQDGVTTIANNLSTVQVFSDEYOLPYVLGSAHQCLPEPPADVEMIPQY 376
QY 241 YLTNNQSAQVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRLAMP 300
DB 377 YLTNNQSAQVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRLAMP 436
QY 301 LIDQYLYNRTQ-NOGSSAQNKDLFSSRGSPPAGMSVQPKNMLPGCYRQORVSKTND 359
DB 437 LIDQYLYNRTQGTGTTNQSRLLFSQAGPQMSLQARNMLPGCYRQORVSKTND 496
QY 360 NNSFTWTGASKYNLNGRSLINPGTAMASHKDEKFFPMGVMIGKESAGASNTALD 419
DB 497 NNSFTWTGASKYNLNGRSLINPGTAMASHKDEKFFPMGVMIGKESAGASNTALD 556
QY 420 NMTTDEEBEIKATNPVATERFGTVAVNFOSSTDPAATGVHAGALPGMWODRDVYLOQ 479
DB 557 NMTTDEEBEIKATNPVATERFGTVAVNFOSSTDPAATGVHAGALPGMWODRDVYLOQ 616
QY 480 PIAKIPHTDGHFHPSPPLMGFGKLPKPPQIMIKNTVPANPPTTFSPAKFASFTQYST 539
DB 617 PIAKIPHTDGHFHPSPPLMGFGKLPKPPQIMIKNTVPANPPTTFSPAKFASFTQYST 676
QY 540 GQVSVIEIEMELQENSKRNMPVEQYTSNYSKASAVDFTVNNGLYTERPRIGRYLTRPL 599
DB 677 GQVSVIEIEMELQENSKRNMPVEQYTSNYSKASAVDFTVNNGLYTERPRIGRYLTRPL 736

RESULT 4

065311

ID 065311 PRELIMINARY; PRT; 736 AA.

AC 065311; 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 3.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=46350;
RN NCBI_TaxID=46350;
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RX MEDLINE=86266430; PubMed=8661429;
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
RT "Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3.";
RT Virology 221:208-217 (1996).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=3H;
RA Muramatsu S., Brown K.E.;
RT Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U48704; AAC55049.1; -
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81660 MW; AFPIEF4B5C67A10 CRC64;

Query Match 86.6%; Score 2815; DB 12; Length 736;
Best Local Similarity 85.2%; Pred. No. 1.8e-193;
Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKRLNFGQTGDSSEVDPQPLGEPATPAA 60
DB 138 TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKRLNFGQTGDSSEVDPQPLGEPATPAA 197
QY 61 VGPTTASGGAPADNNEGADVGNAAGNHCSTMLGDRVITTSRTNALPYYNNHLY 120
DB 198 LGSNTTASGGAPADNNEGADVGNAAGNHCSTMLGDRVITTSRTNALPYYNNHLY 257
QY 121 KQISSASTGASNDNHYFGYSTPWGYFDNRFHCFSPRDMORLINNMGFPRKLSFKLF 180

DB 258 KQISSAS-GASNDNHYFGYSTPWGYFDNRFHCFSPRDMORLINNMGFPRKLSFKLF 316
QY 181 NIOKVEYTNQDGVTTIANNLSTVQVFSDEYOLPYVLGSAHQCLPEPPADVEMIPQY 240
DB 317 NIOKVEYTNQDGVTTIANNLSTVQVFSDEYOLPYVLGSAHQCLPEPPADVEMIPQY 376
QY 241 YLTNNQSAQVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRLAMP 300
DB 377 YLTNNQSAQVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRLAMP 436
QY 301 LIDQYLYNRTQ-NOGSSAQNKDLFSSRGSPPAGMSVQPKNMLPGCYRQORVSKTND 359
DB 437 LIDQYLYNRTQGTGTTNQSRLLFSQAGPQMSLQARNMLPGCYRQORVSKTND 496
QY 360 NNSFTWTGASKYNLNGRSLINPGTAMASHKDEKFFPMGVMIGKESAGASNTALD 419
DB 497 NNSFTWTGASKYNLNGRSLINPGTAMASHKDEKFFPMGVMIGKESAGASNTALD 556
QY 420 NMTTDEEBEIKATNPVATERFGTVAVNFOSSTDPAATGVHAGALPGMWODRDVYLOQ 479
DB 557 NMTTDEEBEIKATNPVATERFGTVAVNFOSSTDPAATGVHAGALPGMWODRDVYLOQ 616
QY 480 PIAKIPHTDGHFHPSPPLMGFGKLPKPPQIMIKNTVPANPPTTFSPAKFASFTQYST 539
DB 617 PIAKIPHTDGHFHPSPPLMGFGKLPKPPQIMIKNTVPANPPTTFSPAKFASFTQYST 676
QY 540 GQVSVIEIEMELQENSKRNMPVEQYTSNYSKASAVDFTVNNGLYTERPRIGRYLTRPL 599
DB 677 GQVSVIEIEMELQENSKRNMPVEQYTSNYSKASAVDFTVNNGLYTERPRIGRYLTRPL 736

RESULT 5

ID 06652 PRELIMINARY; PRT; 735 AA.

AC 06652; 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Major coat protein VP1.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=10804;
RN NCBI_TaxID=10804;
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif.";
RT J. Gen. Virol. 75:0-0(0).
RN [2]
RN SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
DB Dull T., Hoyer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RT Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF043303; AAC03780.1; -
DR InterPro: IPR001403; Parvo_coat.
DR Pfam: PF00740; Parvo_coat; 1.
KM Coat protein.
FT VARIANT 76 D -> V.
FT VARIANT 553 D -> N.
FT VARIANT 567 T -> N.
FT VARIANT 677 QV -> HV.
FT VARIANT 710 V -> R.
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.0%; Score 2764.5; DB 12; Length 735;
Best Local Similarity 82.8%; Pred. No. 7.3e-190;
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKRLNFGQTGDSSEVDPQPLGEPATPAA 60

Db 138 TAPGKREVEHSVPEPDSSSGTGKAGQOPARKRLNFQGTGDADSVDPDQPLGPPAASG 197
 Qy 61 VGPPTMASGGAPADNNEGADGVGNASGNHCHSTLGDVITTSRTTALPTYYNHLX 120
 Db 198 LGITMTAGSGAPADNNEGADGVGNSSGNHCHSTMGDVTTSRTTALPTYYNHLX 257
 Qy 121 KOISSASTGANDNHYFGYSTPMGYPDNFRHCHFSPRDMORLINNMGFPKRLNFKL 180
 Db 258 KOISSQS-GASNDNHYFGYSTPMGYPDNFRHCHFSPRDMORLINNMGFPKRLNFKL 316
 Qy 181 NIQVEVTTNDGVTIANNTSTVOVPSDSEYQLPYVLGSAHQCLPPPADVFMIPQY 240
 Db 317 NIQVEVTTNDGVTIANNTSTVOVPSDSEYQLPYVLGSAHQCLPPPADVFMIPQY 376
 Qy 241 YLTINNGSQAAGRSPFCLEYFSPQMLRTGNFTFSYFEEVPPHSSVAHSQSLDRLNPL 300
 Db 377 YLTINNGSQAAGRSPFCLEYFSPQMLRTGNFTFSYFEEVPPHSSVAHSQSLDRLNPL 436
 Qy 301 LIDQVLYLNTONQSGSAQNKDLLFSRGSPPAGMSVOPKMWLPQPCYRQGRVSKTIDNN 360
 Db 437 LIDQVLYLNTONQSGSAQNKDLLFSRGSPPAGMSVOPKMWLPQPCYRQGRVSKTIDNN 496
 Qy 361 NSNFTWGTASRYNLNGRBSIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDN 420
 Db 497 NSEYMTGATKYHLNGRBSIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDN 556
 Qy 421 VMIDDEERIKATNPVAREFCTVAVNFQSSSTDPATGDVHMGALPGVMWODRVPYLOGP 480
 Db 557 VMIDDEERIKATNPVAREFCTVAVNFQSSSTDPATGDVHMGALPGVMWODRVPYLOGP 616
 Qy 481 IMAKIPHTDGHFHSPLMGFGGLKNPPOILIKNTPVPANPAPASATKFASTTOYSTG 540
 Db 617 IMAKIPHTDGHFHSPLMGFGGLKNPPOILIKNTPVPANPAPASATKFASTTOYSTG 676
 Qy 541 QVSVEIEMWELOKENSKRNPEVQTSNYSKASANDFTVDNNGLTERRPIGTRVLTPEL 599
 Db 677 QVSVEIEMWELOKENSKRNPEVQTSNYSKASANDFTVDNNGLTERRPIGTRVLTPEL 735

RESULT 6
 056653 PRELIMINARY; PRT; 598 AA.
 AC 056653;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Major coat protein VP2.
 OS Adeno-associated virus 2 (AAV2).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=10804;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95088582; PubMed=7996133;
 RA Ruffing M., Heid H., Kleinschmidt J.A.,
 RA "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
 RT proteins affect viral infectivity: lack of an RGD integrin-binding
 RT motif";
 RT J. Gen. Virol. 75:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
 RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
 RA Tratschin J.-D., Weitz M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF043303; AAC03778.1;
 DR InterPro; IPR001403; Parvo_coat; 1.
 DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein.
 FT VARIANT 416 416 D -> N.
 FT VARIANT 430 430 T -> N.
 FT VARIANT 540 541 QV -> HV.
 FT VARIANT 573 573 V -> R.
 SQ SEQUENCE 598 AA; 66619 MM; 070811ED9368B934 CRC64;

Query Match 84.9%; Score 2759.5; DB 12; Length 598;
 Best Local Similarity 82.8%; Pred. No. 1.3e-189;
 Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

Qy 2 APGKREVEHSVPEPDSSSGTGKAGQOPARKRLNFQGTGDADSVDPDQPLGPPAASG 61
 Db 2 APGKREVEHSVPEPDSSSGTGKAGQOPARKRLNFQGTGDADSVDPDQPLGPPAASG 61
 Qy 62 GPTTMSGGAPADNNEGADGVGNASGNHCHSTLGDVITTSRTTALPTYYNHLX 121
 Db 62 GPTTMSGGAPADNNEGADGVGNASGNHCHSTLGDVITTSRTTALPTYYNHLX 121
 Qy 122 QISSASTGANDNHYFGYSTPMGYPDNFRHCHFSPRDMORLINNMGFPKRLNFKL 181
 Db 122 QISSQS-GASNDNHYFGYSTPMGYPDNFRHCHFSPRDMORLINNMGFPKRLNFKL 180
 Qy 181 NIQVEVTTNDGVTIANNTSTVOVPSDSEYQLPYVLGSAHQCLPPPADVFMIPQY 240
 Db 181 NIQVEVTTNDGVTIANNTSTVOVPSDSEYQLPYVLGSAHQCLPPPADVFMIPQY 240
 Qy 241 YLTINNGSQAAGRSPFCLEYFSPQMLRTGNFTFSYFEEVPPHSSVAHSQSLDRLNPL 300
 Db 241 YLTINNGSQAAGRSPFCLEYFSPQMLRTGNFTFSYFEEVPPHSSVAHSQSLDRLNPL 300
 Qy 301 LIDQVLYLNTONQSGSAQNKDLLFSRGSPPAGMSVOPKMWLPQPCYRQGRVSKTIDNN 361
 Db 301 LIDQVLYLNTONQSGSAQNKDLLFSRGSPPAGMSVOPKMWLPQPCYRQGRVSKTIDNN 360
 Qy 361 NSNFTWGTASRYNLNGRBSIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDN 420
 Db 361 NSNFTWGTASRYNLNGRBSIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDN 420
 Qy 421 VMIDDEERIKATNPVAREFCTVAVNFQSSSTDPATGDVHMGALPGVMWODRVPYLOGP 481
 Db 421 VMIDDEERIKATNPVAREFCTVAVNFQSSSTDPATGDVHMGALPGVMWODRVPYLOGP 480
 Qy 481 IMAKIPHTDGHFHSPLMGFGGLKNPPOILIKNTPVPANPAPASATKFASTTOYSTG 541
 Db 481 IMAKIPHTDGHFHSPLMGFGGLKNPPOILIKNTPVPANPAPASATKFASTTOYSTG 540
 Qy 541 QVSVEIEMWELOKENSKRNPEVQTSNYSKASANDFTVDNNGLTERRPIGTRVLTPEL 599
 Db 541 QVSVEIEMWELOKENSKRNPEVQTSNYSKASANDFTVDNNGLTERRPIGTRVLTPEL 598

RESULT 7
 092917 PRELIMINARY; PRT; 533 AA.
 AC 092917;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Major coat protein VP3.
 OS Adeno-associated virus 2 (AAV2).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=10804;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95088582; PubMed=7996133;
 RA Ruffing M., Heid H., Kleinschmidt J.A.,
 RA "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
 RT proteins affect viral infectivity: lack of an RGD integrin-binding
 RT motif";
 RT J. Gen. Virol. 75:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
 RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
 RA Tratschin J.-D., Weitz M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF043303; AAC03779.1;
 DR InterPro; IPR001403; Parvo_coat;

DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein. 351 D -> N.
 FT VARIANT 365 T -> N.
 FT VARIANT 475 QV -> HV.
 FT VARIANT 508 V -> R.
 SQ SEQUENCE 533 AA; 60063 MM; 9E4DBBC25810D4F0 CRC64;

Query Match 76.3%; Score 2481.5; DB 12; Length 533;
 Best Local Similarity 83.3%; Pred. No. 9.2e-170;
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

66 MASGCGAPMADNNGAGDVGNAAGNMGHCDSTWLGDRYITTSRTWALPTYNHLYKXISS 125
 1 MATSGAPMADNNEGAGVGNSGNGHCDSTWMDRITTSRTWALPTYNHLYKXISS 60
 126 ASTGASNDNHYFGYSTPWGDFDNRPHCHSPRDMQRLINNMGFRPKRLNFKLFINIQVK 185
 61 QS-GASNDNHYFGYSTPWGDFDNRPHCHSPRDMQRLINNMGFRPKRLNFKLFINIQVK 119
 186 EVTTNDGVTITANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPADVPMIPOYGYLTIN 245
 120 EVTQNDGVTITANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVPMIPOYGYLTIN 179
 246 NGSQAVGRSSFFCYCLEYFPSCWMLRTGNNTFSYFEEVPHSSVAHSQSLDLNMLIDQY 305
 180 NGSQAVGRSSFFCYCLEYFPSCWMLRTGNNTFSYFEEVPHSSVAHSQSLDLNMLIDQY 239
 306 LYYLNRITQNGSAGNKKDLFSRGSFAGMSVOPKMLPGPCYRQQRVSKTKTDNNNSFT 365
 240 LYYLNRITQNGSAGNKKDLFSRGSFAGMSVOPKMLPGPCYRQQRVSKTKTDNNNSFT 299
 366 WTGASKNLNGBREIINPTGAMASHKDDKFFPMGSMVIFPKESAGASATLADNMTD 425
 300 WTGATKXHLNGRSLVNPFGPAMASHKDDKFFPMGSMVIFPKESAGASATLADNMTD 359
 426 EEEIKATNPVATERFGVAVNFQSSSTDPATGVHAGALPGMWQORVDYLGQPIYAKI 465
 360 EEEIKATNPVATERFGVAVNFQSSSTDPATGVHAGALPGMWQORVDYLGQPIYAKI 419
 486 PHTDGHFHPPLMGFGFLKNPPQILIKNTPVPANPAEFSAFTQYSTGVSV 545
 420 PHTDGHFHPPLMGFGFLKNPPQILIKNTPVPANPAEFSAFTQYSTGVSV 479
 546 IEMELQKENSKRNPPEVQYTSNYSKASANDVTUNNGLYTEPRIGRILYTRPL 599
 480 IEMELQKENSKRNPPEVQYTSNYSKASANDVTUNNGLYTEPRIGRILYTRPL 533

RESULT 8

041855 PRELIMINARY; PRT; 734 AA.
 AC 041855;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Capsid.
 OS adeno-associated virus 4.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=57579;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97404695; PubMed=9261407;
 RX Clontini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;
 RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of
 recombinant AAV4 particles."
 RL J. Virol. 71:6823-6833(1997).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC VR-646;
 RA Clontini J.A., Yang L., Kotin R.M., Safer B.;
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; U89790; AAC58045.1; -.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 SQ SEQUENCE 734 AA; 80639 MM; 616CC27A777BBE6F CRC64;

Query Match 56.3%; Score 1830.5; DB 12; Length 734;
 Best Local Similarity 57.8%; Pred. No. 7e-123;
 Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;

1 TAFGKRPEVQSPQEPDSSGIGKTQQAQKRLNFGQTSSESVDDPQICEPATPA 59
 137 TAFGKRPLESPQEPDSSGIGKTQQAQKRLNFGQTSSESVDDPQICEPATPA 190
 60 AVPEPTMASGCGAPMADNNEGAGVGNAAGNMGHCDSTWLGDRYITTSRTWALPTYNHLY 119
 191 MSDSEMRALAAAGAAVEGGAGDVGNAAGNMGHCDSTWEGHVTITSTRTWALPTYNHLY 250
 120 YKQISSASTGASNDNHYFGYSTPWGDFDNRPHCHSPRDMQRLINNMGFRPKRLNFKL 179
 251 YKRLGE-----SLQSTNYNGSTPWGDFDNRPHCHSPRDMQRLINNMGFRPKRLNFKL 306
 180 FNIQVKEVTTNDGVTITANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPADVPMIPOY 239
 307 FNIQVKEVTTNDGVTITANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPADVPMIPOY 366
 240 GY---LTINNGSAGVGRSSFFCYCLEYFPSCWMLRTGNNTFSYFEEVPHSSVAHSQSLDR 236
 367 GYGLVATGNTSQQDTDRNMFYCLEYFPSCWMLRTGNNTFSYFEEVPHSSVAHSQSLDR 426
 297 LAMPPLIDQYLYNTRTON-----QSGAGNKKDLFSRGSFAGMSVOPKMLPGPCYRQQRV 352
 427 LAMPPLIDQYLYNTRTON-----QSGAGNKKDLFSRGSFAGMSVOPKMLPGPCYRQQRV 483
 353 SKTKTDNNNSNFTWTGAS---KY---NINRESIINPTGAMASHKDDKFFPMGSMVIF 405
 484 SKTA---NQYKIPATGSDSLIKETSTLDRGMSALTPGPMAIAGPADSK-FSNQLIF 540
 406 FPKESAGASTLADNMTDEEIKATNPVATERFGVAVNFQSSSTDPATGVHAGALPGMW 465
 541 AGPKONGNTATVPDGTULFTSEELATINATIDTDMWNLPGQDSSNLPFLVDRLTLGAV 600
 466 PGWQORVDYLGQPIYAKIPHTDGHFHPPLMGFGFLKNPPQILIKNTPVPANPAEF 525
 601 PGWQORVDYLGQPIYAKIPHTDGHFHPPLMGFGFGFLKNPPQILIKNTPVPANPAEF 660
 526 SATKFSFTQYSTGVSVIEMELQKENSKRNPPEVQYTSNYSKASANDVTUNNGLYT 565
 661 SSTPVNSFTQYSTGVSVQIDWEIQERSKRNPPEVQYTSNYSKASANDVTUNNGLYT 720
 586 EPRIGTRILYTRPL 599
 721 EPRIGTRILYTRPL 734

RESULT 9

067666 PRELIMINARY; PRT; 732 AA.
 AC 067666;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE VP1.
 OS goose parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=8251;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VIRULENT B;
 RX MEDLINE=96010229; PubMed=7571426;
 RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
 RT "Analysis of the complete nucleotide sequences of goose and muscovy
 duck parvoviruses indicates common ancestral origin with adeno-

RT associated virus 2.":
RL Virology 212:562-573 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25749; AAA83230.1; --
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC54;
Query Match 52.8%; Score 1717; DB 12; Length 732;
Best Local Similarity 53.0%; Pred. No. 9.6e-115; Mismatches 148; Indels 48; Gaps 13;
Matches 327; Conservative
QY 1 TAPKKR-----PVEQSPQ-EPDSSSGIGKTGQOPAKKRLNFGQTDSEVPDPL 51
Db 146 TAPAKKNTGKLTIDHYPVYKPKLTBEVAGGSSAVDQ-----GATAB----- 189
QY 52 GEPATPAVPTTMAAGGAPMADNNEGADGVNAGSNMHCDSITWLDGRTVITSTRTWA 111
Db 190 GTEP-----VAASEMAEGGAGMDSSGADGVNAGSNMHCDSQMMGNVITTKTRTW 244
QY 112 LPTNNHLKYKISSASTGASNDNH--YFGSTPMGYPDFNRFCHFSFPRDQRLINNMW 169
Db 245 LPTNNHLKYKITS---GTSODANVQAGYSTPMGYDFNRFCHFSFPRDQRLINNMW 301
QY 170 FRPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFSSEYOLPYVLSAHQCLPPF 229
Db 302 IRPKSLKFIKNVQVKEVTTODQKTIANNLTSTIQVFTDDEHQLPYVLSATGTMPPF 361
QY 230 PADFMFPQVGYTLN---NGSQAVGSSFFCYCLEYFPQMLRTGNFTFSTFEVPHS 286
Db 362 PSDYVALPQVGYCTMHTNQGARNDRSAFYCLEYFPQMLRTGNFTFSTFEVPHS 421
QY 287 SYASQSLDRMLNPLIDQYLYLNRTQNGSAQNKLLFSRGSPPAGMSVQPKMLPGPC 346
Db 422 MFAISQDLRLMPLVQYLYLNRTQNGSAQNKLLFSRGSPPAGMSVQPKMLPGPC 475
QY 347 YRQGRV-SKTKTDNNNSNFT-WTGASKYLNNGRESIINPTAMASHKODEKFFPMGVM 404
Db 476 FLDRVAVRAYTGNTNANMINSNGNKVNLKDRQYLQPPVSAITYEGEASSLPANIL 535
QY 405 IFGE--SAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPADGPHAM 462
Db 536 GIAKDPYRSGSTTAGISDIMWTEOEVAPTNGVGKPYGTVTNEQNTTAPATSSDLVL 595
QY 463 GALEGMYQDRDVLQGPIMAKIPHTDGHFHPSPMLGFGGLNPPQILLKNTFVPANP 522
Db 596 GALEGMYQDRDVLQGPIMAKIPHTDGHFHPSPMLGFGGLNPPQILLKNTFVPANP 655
QY 523 AEFSAITFASITTOYSTGQVSEIEMELQKNSKRNPEVOYTSNYAKSANDVTVDNNG 582
Db 656 VEYHOKNSYITTOYSTGQCTEVMWELRKENSXKRNPEIQFISNFSNRSTIMPAEMETG 715
QY 583 LYTEPRITGRYTLRPL 599
Db 716 GYVEDRLIGRYTLQNL 732

RESULT 10
083290 PRELIMINARY; PRT; 732 AA.
AC 083290;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Capsid protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI TaxID=37325;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=89384/France;
RX MEDLINE=96406928; PubMed=8811015;
RA Le Gall-Reculle G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
RT Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)
in a baculovirus expression system and demonstration of immunity
induced by the recombinant protein.";
RL J. Gen. Virol. 77:2159-2163 (1996).
DR EMBL; Z68272; CA92575.1; --
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
FT CHAIN 146 732 VP2 CAPSID PROTEIN.
FT CHAIN 199 732 VP3 CAPSID PROTEIN.
SQ SEQUENCE 732 AA; 81364 MW; DE70DCB215F4E2 CRC64;
Query Match 52.7%; Score 1714; DB 12; Length 732;
Best Local Similarity 53.3%; Pred. No. 1.6e-114; Mismatches 164; Indels 40; Gaps 11;
Matches 327; Conservative
QY 6 KRPEVSGPQEBDSSSGIGK-TGQOPAKKRLNFGQTDSEVPDPLG-----E 53
Db 141 EEPVNTAPAKSS-----GKLTIDHYPVYKPKLSE---ENSPSPNSGGASAAATGSE 192
QY 54 PPATPAVPTTMAAGGAPMADNNEGADGVNAGSNMHCDSITWLDGRTVITSTRTWA 113
Db 193 PVAP-----NMAEGSGAGMDSSGADGVNAGSNMHCDSQWLGDTVITTKTRTW 246
QY 114 TNNHLKYKISSASTGASNDNH--YFGSTPMGYPDFNRFCHFSFPRDQRLINNMW 173
Db 247 SYNHHLYKALTSQTNPDNS--TQYAGYSTPMGYDFNRFCHFSFPRDQRLINNMW 305
QY 174 RLNFKLNIQVKEVTNDGVTIANNLSTVQVFSSEYOLPYVLSAHQCLPPPADY 233
Db 306 ALFKIFPNVQVKEVTTODQKTIANNLTSTIQVFTDDEHQLPYVLSATGTMPPFSDV 365
QY 234 FMIPQGYTLN---NGSQAVGSSFFCYCLEYFPQMLRTGNFTFSTFEVPHS 290
Db 366 YALPQGYCTMHTNQGARNDRSAFYCLEYFPQMLRTGNFTFSTFEVPHS 425
QY 291 SOSLDRLMPLIDQYLYLNRTQNGSAQNKLLFSRGSPPAGMSVQPKMLPGPC 350
Db 426 SOSLDRLMPLIDQYLYLNRTQNGSAQNKLLFSRGSPPAGMSVQPKMLPGPC 479
QY 351 RVSKTK--TDNNNSNFTWTASKYLNNGRESIINPTAMASHKODEKFFPMGVM 408
Db 480 RVKAYSGDTNANMINSNGNKVNLKDRQYLQPPVSAITYEGEASSLPANIL 539
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERGTVAVNFQSSSTDPADGPHAM 466
Db 540 DPYRSGSTTAGISDIMWTEOEIAPNTNGVGMRYGLVTNEQNTTAPATSSDLVL 599
QY 467 GMYQDRDVLQGPIMAKIPHTDGHFHPSPMLGFGGLNPPQILLKNTFVPANP 526
Db 600 GMYQDRDVLQGPIMAKIPHTDGHFHPSPMLGFGGLNPPQILLKNTFVPANP 659
QY 527 ATEFASITTOYSTGQVSEIEMELQKNSKRNPEVOYTSNYAKSANDVTVDNNG 586
Db 660 NQKNSYITTOYSTGQCTEVMWELRKENSXKRNPEIQFISNFSNRSTIMPAEMETG 719
QY 587 PRPIGRYTLRPL 599
Db 720 DRLIGRYTLQNL 732

RESULT 11
067667 PRELIMINARY; PRT; 587 AA.
AC 067667;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VP2.
OS VP2.
OC
NCBI TaxID=37325;
RN [1]

OS goose parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=38251;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VIRULENT B;
 RX MEDLINE=96010229; PubMed=7571426;
 RA Zadori Z., Stefancsik R., Rauch T., Kisary J.
 RT "Analysis of the complete nucleotide sequences of goose and muscovy
 duck parvoviruses indicates common ancestral origin with adeno-
 RT associated virus 2."
 RN Virology 212:562-573 (1995).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=VIRULENT B;
 RA Zadori Z.
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U25743; AA83231.1; -
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 SQ SEQUENCE 587 AA; 65246 MW; 02785D5FE7F0423 CRC64;

Query Match 52.7%; Score 1712; DB 12; Length 587;
 Best Local Similarity 52.9%; Pred. No. 1.6e-114;
 Matches 326; Conservative 94; Mismatches 148; Indels 48; Gaps 13;

QY 2 APGKKR-----PVEQSPQ-EPDSSSGIGTGGQPAKKRLNPGQTGDSSEVPDPPLG 52
 DB 2 APAKNTGKLTIDHYVVKPKRLTEVSAGGSSAVQDG-----GATAE-----G 45
 QY 53 EPPATPAAVGPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTWA 112
 DB 46 TEP-----VAASEMAEGGGAGMGDSGGADGVGNAGNWHCDSTWLGDRVITSTRTWA 100
 QY 113 PTYNNHLYKQISSASTASNDN--YFGYSTPWGTFDFFNRFCFSPDQRLINNMWG 170
 DB 101 PSYNNHLYKAIT--GTSQDANVQAGYSTPWGTFDFFNRFCFSPDQRLINNMWG 157
 QY 171 RPKRLNFKLNFIOVKEVTNDOGTIANNLTSTVQVFSDESEYOLPYVLGSAHQGLPP 230
 DB 158 RPKSLKFKIRNVOVKEVTNDOGTIANNLTSTVQVFSDESEYOLPYVLGSAHQGLPP 217
 QY 231 ADVFMIPOGYLTLLN--NGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHS 287
 DB 218 SDVYALPOGYCTMHTNNGARFNDRSAFYCLEYFSPQMLRTGNFTFSYTFEEVPHS 277
 QY 288 YAHQSOLRLNPLIDQYLYLNTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCY 347
 DB 278 FAHSODLRLNPLIDQYLYLNTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCY 331
 QY 348 RQGRV-SKTKTDNNNSNFT-WTGASKYLNNGRESIIINPGTMAASHKDEDEKFFMSGVM 405
 DB 332 LDRVRVRYTGTDVYANNMINSNGKYNLKRQYLLQPGVSAATYVEGEASSLPQNILL 391
 QY 406 FGKE--SAGASNTALDNMTTDEEIKATNPVATERGTVAANFOSSSTPATGDVHAM 463
 DB 392 IADPRSSGTTAGISDINWTEDEVAFTNGVCKPGRVTNBNQNTTAPSSDLDVL 451
 QY 464 ALPGMWQDRDVIYLGPIWAKIPHTDGFHPSPLMGFGGLKNPPQILIKNTVPANPA 523
 DB 452 ALPGMWQDRDVIYLGPIWAKIPHTDGFHPSPLMGFGGLKNPPQILIKNTVPANPA 511
 QY 524 EFAATKFASTIYQSTGVSVIELELOKENSKRNNPEVQYTSNKAASAVDFYVNDNG 563
 DB 512 EYVHQKNSYITQYSTGVSVIELELOKENSKRNNPEVQYTSNKAASAVDFYVNDNG 571
 QY 584 YTEPRPIGRTYLTPL 599
 DB 572 YVEDRLIGRTYLTPL 587

RESULT 12
 Q8V395

ID Q8V395 PRELIMINARY; PRT; 732 AA.
 AC Q8V395;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Capsid protein VP.
 GN VP.
 OS goose parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=38251;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GPV-YG;
 RA Ge Y., You Y., Xu Q.
 RT "Analysis of the major open reading frames' nucleotide sequences in
 RT Goose parvovirus GPV-YG strain isolated in China."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF416726; AA137722.1; -
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 SQ SEQUENCE 732 AA; 81456 MW; 73F2E4BC769744B6 CRC64;

Query Match 52.7%; Score 1712; DB 12; Length 732;
 Best Local Similarity 53.2%; Pred. No. 2.2e-114;
 Matches 328; Conservative 89; Mismatches 152; Indels 48; Gaps 13;

QY 1 TAPGKKR-----PVEQSPQ-EPDSSSGIGTGGQPAKKRLNPGQTGDSSEVPDPPL 51
 DB 146 TAPKNTGKLTIDHYVVKPKRLTEVSAGGSSAVQDG-----GATAE----- 189
 QY 52 EPPATPAAVGPTTMAAGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTWA 111
 DB 190 GTEP-----VAASEMAEGGGAGMGDSGGADGVGNAGNWHCDSTWLGDRVITSTRTWA 244
 QY 112 LPTNNHLYKQISSASTASNDN--HYFGYSTPWGTFDFFNRFCFSPDQRLINNMWG 169
 DB 245 LPSYNNHLYKAIT--GTSQDANVQAGYSTPWGTFDFFNRFCFSPDQRLINNMWG 301
 QY 170 RPKRLNFKLNFIOVKEVTNDOGTIANNLTSTVQVFSDESEYOLPYVLGSAHQGLPP 229
 DB 302 RPKSLKFKIRNVOVKEVTNDOGTIANNLTSTVQVFSDESEYOLPYVLGSAHQGLPP 361
 QY 230 PADVMIPOGYLTLLN--NGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHS 286
 DB 362 PSDVYALPOGYCTMHTNNGARFNDRSAFYCLEYFSPQMLRTGNFTFSYTFEEVPHS 421
 QY 287 YAHQSOLRLNPLIDQYLYLNTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPC 346
 DB 422 MFHSODLRLNPLIDQYLYLNTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPC 475
 QY 347 YRQGRV-SKTKTDNNNSNFT-WTGASKYLNNGRESIIINPGTMAASHKDEDEKFFMSGVM 404
 DB 476 LDRVRVRYTGTDVYANNMINSNGKYNLKRQYLLQPGVSAATYVEGEASSLPQNILL 535
 QY 405 IFKE--SAGASNTALDNMTTDEEIKATNPVATERGTVAANFOSSSTPATGDVHAM 462
 DB 536 GLAKDYRSGTSTAGISDINWTEDEVAFTNGVCKPGRVTNBNQNTTAPSSDLDVL 565
 QY 463 GALPGMWQDRDVIYLGPIWAKIPHTDGFHPSPLMGFGGLKNPPQILIKNTVPANPA 522
 DB 596 GALPGMWQDRDVIYLGPIWAKIPHTDGFHPSPLMGFGGLKNPPQILIKNTVPANPA 655
 QY 523 AEFSAKFASTIYQSTGVSVIELELOKENSKRNNPEVQYTSNKAASAVDFYVNDNG 562
 DB 656 VEYVHQKNSYITQYSTGVSVIELELOKENSKRNNPEVQYTSNKAASAVDFYVNDNG 715
 QY 583 LYTEPRPIGRTYLTPL 599
 DB 716 GYIEDRLIGRTYLTPL 732

RESULT 13
 Q65444

ID 065444 PRELIMINARY; PRT; 732 AA.
AC 065444;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Capsid protein VP.
GN VP.
OS Barbary duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=39118;
RX STRAIN=FM;
RP SEQUENCE FROM N.A.
RX MEDLINE=96010229; PubMed=7571246;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573(1995).
DR EMBL; U22967; AAA83225.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81314 MW; D92360596E2D2C05 CRC64;

Query Match 52.4%; Score 1702; DB 12; Length 732;
Best Local Similarity 53.0%; Pred. No. 1,4e-113; Indels 40; Gaps 11;
Matches 325; Conservative 83; Mismatches 165;

QY 6 KRPEQSGPOBPDSSSGIGK-TGQOPAKKRLNFGQTDGSEVPDQPLG-----E 53
DB 141 EEPVNTAPAKKS-----GKLTDDHPYVKKPKLSE-----ENSPSPNSGSGASAAATGSE 192
QY 54 PPATPAVGPPTMASGGAGPMAADNNEGADGVGNASGNMHCDSQWLDGTVITKTRTWLP 113
DB 193 FVAAP-----NMAEGSGAGMDSAGADGVGNASGNMHCDSQWLDGTVITKTRTWLP 246
QY 114 TYNHLYKQISSASTGASNDNHFGYSTPMGYDFNFRFCHFSRPDQRLINNMGRPK 173
DB 247 SYNHHMQLTSGTNPDSN-TQYAGYSTPMGYDFNFRFCHFSRPDQRLINNMGRPK 305
QY 174 RLNFKLFNIOVEKVTNDGVTIANNLSTVQVFSSEYOLPYVLSAHQGLPPPADV 233
DB 306 ALKFKEFNVQKVTYTDQDKTIANNLSTIQIFTDNEHQLPYVLSAHQGLPPPADV 365
QY 234 FMIPQGYLTN---NGSQAVGRSSFYCLEYFPSSQMLRTGNFTFSYTEEVPFHSYAH 290
DB 366 YALPQGYCTMHTNQGAFNDRSAPFYCLEYFPSSQMLRTGNFTFSYTEEVPFHSYAH 425
QY 291 SQSLDRMLNPILDIQYLYLNRTQNOGSAQNKDLLFSRGSPPAGMSVQPKMLPGPCYRQ 350
DB 426 SQDLDRMLNPILDIQYLYLNRTQNOGSAQNKDLLFSRGSPPAGMSVQPKMLPGPCYRQ 479
QY 351 RVSKTK--TDNNNSNFTWTGASKYNLNGRESIINPGTAMASEKDEDEKFFPMGVMIFGK 408
DB 480 RVRAVSGGTDNYANMSIWSKGNKVFLLKREYLLQPGVATHTHEDQASSVPAQNIIGIAK 539
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATDVAHMGALP 466
DB 540 DPYRSGSTLAGISDIWTDQELAPNTGVGRPYGLTVNBEQTTAPTALEVLGALP 599
QY 467 GMMWQBDVYLOQPIAKIPHTDGHFHPSPMLGFGGLKNPPQILIKNTVPVAPNPAEFS 526
DB 600 GMMWQNRDIYLOQPIAKIPKTDGKPHSPMLGFGGLKNPPQILIKNTVPVAPNPAEVL 659
QY 527 ATKFASFTIYQSTGVSEIELELQKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTE 586
DB 660 NQKMSYITQYSTGQCTVEVMELELQKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTE 719
QY 587 PRPIGRTYLRPL 599
DB 720 DRLIGRTYLRPL 732

RESULT 14
ID 083289 PRELIMINARY; PRT; 732 AA.
AC 083289;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Vp1 protein.
GN Vp1.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=37325;
RX STRAIN=FM;
RP SEQUENCE FROM N.A.
RA Zadori Z., Erdei J., Nagy J., Kisari J.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X75093; CA52984.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81284 MW; 9FF6955BC15DF3A CRC64;

Query Match 52.3%; Score 1701; DB 12; Length 732;
Best Local Similarity 53.0%; Pred. No. 1,4e-113; Indels 40; Gaps 11;
Matches 325; Conservative 83; Mismatches 165;

QY 6 KRPEQSGPOBPDSSSGIGK-TGQOPAKKRLNFGQTDGSEVPDQPLG-----E 53
DB 141 EEPVNTAPAKKS-----GKLTDDHPYVKKPKLSE-----ENSPSPNSGSGASAAATGSE 192
QY 54 PPATPAVGPPTMASGGAGPMAADNNEGADGVGNASGNMHCDSQWLDGTVITKTRTWLP 113
DB 193 FVAAP-----NMAEGSGAGMDSAGADGVGNASGNMHCDSQWLDGTVITKTRTWLP 246
QY 114 TYNHLYKQISSASTGASNDNHFGYSTPMGYDFNFRFCHFSRPDQRLINNMGRPK 173
DB 247 SYNHHMQLTSGTNPDSN-TQYAGYSTPMGYDFNFRFCHFSRPDQRLINNMGRPK 305
QY 174 RLNFKLFNIOVEKVTNDGVTIANNLSTVQVFSSEYOLPYVLSAHQGLPPPADV 233
DB 306 ALKFKEFNVQKVTYTDQDKTIANNLSTIQIFTDNEHQLPYVLSAHQGLPPPADV 365
QY 234 FMIPQGYLTN---NGSQAVGRSSFYCLEYFPSSQMLRTGNFTFSYTEEVPFHSYAH 290
DB 366 YALPQGYCTMHTNQGAFNDRSAPFYCLEYFPSSQMLRTGNFTFSYTEEVPFHSYAH 425
QY 291 SQSLDRMLNPILDIQYLYLNRTQNOGSAQNKDLLFSRGSPPAGMSVQPKMLPGPCYRQ 350
DB 426 SQDLDRMLNPILDIQYLYLNRTQNOGSAQNKDLLFSRGSPPAGMSVQPKMLPGPCYRQ 479
QY 351 RVSKTK--TDNNNSNFTWTGASKYNLNGRESIINPGTAMASEKDEDEKFFPMGVMIFGK 408
DB 480 RVRAVSGGTDNYANMSIWSKGNKVFLLKREYLLQPGVATHTHEDQASSVPAQNIIGIAK 539
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATDVAHMGALP 466
DB 540 DPYRSGSTLAGISDIWTDQELAPNTGVGRPYGLTVNBEQTTAPTALEVLGALP 599
QY 467 GMMWQBDVYLOQPIAKIPHTDGHFHPSPMLGFGGLKNPPQILIKNTVPVAPNPAEFS 526
DB 600 GMMWQNRDIYLOQPIAKIPKTDGKPHSPMLGFGGLKNPPQILIKNTVPVAPNPAEVL 659
QY 527 ATKFASFTIYQSTGVSEIELELQKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTE 586
DB 660 NQKMSYITQYSTGQCTVEVMELELQKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTE 719
QY 587 PRPIGRTYLRPL 599
DB 720 DRLIGRTYLRPL 732

RESULT 15
ID 09YIU1 PRELIMINARY; PRT; 724 AA.

AC Q9Y101; 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Capsid protein.
 GN CAP OR VP1.
 OS adeno-associated virus 5.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=82300;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99098980; PubMed=9882294;
 RA Bantel-Schaal U., Delius H., Schmidt R., Zur Hausen H.;
 RT "Human adeno-associated virus type 5 is only distantly related to
 RT other known primate helper-dependent parvoviruses."
 RL J. Virol. 73:939-947(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9909022; PubMed=9882336;
 RA Chlorini J.A., Kim F., Yang L., Kotin R.M.;
 RT "Cloning and characterization of adeno-associated virus type 5."
 RL J. Virol. 73:1309-1319(1999).
 DR EMBL; Y18065; CAJ7024.1; -.
 DR EMBL; AF085716; AAD13756.1; -.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 SQ SEQUENCE 724 AA; 80424 MW; BBC332888258D34C CRC64;

Query Match 52.3%; Score 1700.5; DB 12; Length 724;
 Best Local Similarity 55.0%; Pred. No. 1.4e-113;
 Matches 332; Conservative 64; Mismatches 173; Indels 35; Gaps 7;

QY 3 PGKKRPVEQSPQSPSSSGIGKTGQPAKKRLNFCGTGSESVDPQPLGEPATPA-AV 61
 DB 149 PKKKKARTEDSKPTSS-----DAAGSGSQQLIPAPASSL 188
 QY 62 GPTTMAAGGAPADNNEGADGVNAGSNWCHDSTWLGDRVITSTRTALPTYNNHLYK 121
 DB 189 GADTMSAGGGPLGDNNGADGVNAGSNWCHDSTWLGDRVITSTRTALPTYNNHLYK 248
 QY 122 QISSASTGASNDNHRYGYSYTPWGYFDPNRFCHSPRDQRLINNMGFRPKRLNFKLEN 181
 DB 249 EIKSGSVDSGNANAYFGYSTPWGYPFDRFHSWSPRDQRLINNMGFRPKRLNFKLEN 308
 QY 182 IQVEVTTNDGVTIANNLSTVOVPSDSRYOLPYULGSAHOGCLPPFPADVFMIPQYX 241
 DB 309 IQVEVTVQDSTTTIANNLSTVOVFTDDYQLPYVNGTGEGLPAFPPOVFTLPQYX 368
 QY 242 LTLN--NGSOAVGRSSFFCLEYFPSONLRTGNNFTFSYFEEVFPFSSVYASQSLDLN 299
 DB 369 ATIANRNTENTPRTSSFFCLEYFPSONLRTGNNFTFSYFEEVFPFSSVYASQSLDLN 428
 QY 300 PLIDQYLILYINRTQONSGAQNKLFLSRGSPAGMSVOP-KNMLPGFCYRQQRVSKTKTD 358
 DB 429 PLVDQYLIRFVSTNTGTVQFNKML-----AGRYANTYKNWPGPMGRTQGMNLGSGV 481
 QY 359 NNNNSFTWGSACYNLNGRESIINPGTAMASHKDDKFFPMGCVMI FGKESAGASTAL 418
 DB 482 NRASVSFAFTNTNMELEGASVYVPPQNGMTNNLQGSNTALENTMIFNSOPANPGTTAT 541
 QY 419 ---DNWITDEBEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWODRDV 475
 DB 542 YLEGNMLITSESETPQPNRVAIVVGQMATNNSSTTAPATGYNLQELIPGSVMMERDV 601
 QY 476 YLOGPTWAKIPHTDGHFHPSPLMGFGFLKNPPQILIKNTVPANPPAEFSATKFSFIT 535
 DB 602 YLOGPTWAKIPHTDGHFHPSPAMGFGFLKHPPMWLIKNTVPAN--ITSFSDVPSFIT 660
 QY 536 QYSTGVSVIEMELQKENSRRNNDVQYTSNYSKANSANVDFTDNNGLYTEPRPIGTRYL 595
 DB 661 QYSTGVSVIEMELQKENSRRNNDVQYTSNYSKANSANVDFTDNNGLYTEPRPIGTRYL 720
 QY 596 TRPL 599

DB 721 TRPL 724

Search completed: July 17, 2003, 18:34:31
 Job time: 46.5484 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 : Search time 37.1429 Seconds
(without alignments)
1915.734 Million cell updates/sec

Title: US-09-807-802a-17

Perfect score: 2906

Sequence: 1 MASGGAGPMADNMGADGVG.....NNGLYTEPRPIGTRVLTPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2906	100.0	534	21	AAV71169
2	2906	100.0	599	21	AAV71168
3	2906	100.0	736	21	AAV71167
4	2884	99.2	736	22	AAV59847
5	2530	87.1	736	22	AAV59846
6	2528	87.0	736	22	AAV59845
7	2481.5	85.4	533	22	AAV65794
8	2481.5	85.4	598	22	AAV65793
9	2481.5	85.4	598	22	AAV51509
10	2481.5	85.4	735	22	AAV65792

11	2481.5	85.4	735	22	AAV51508
12	2481.5	85.4	735	22	AAV59844
13	2477.5	85.3	533	22	AAV51510
14	2455	84.5	734	22	AAV50326
15	1690.5	58.2	534	19	AAV46314
16	1690.5	58.2	598	19	AAV46313
17	1690.5	58.2	734	19	AAV46308
18	1690	58.2	736	18	AAV19000
19	1676	57.7	534	16	AAV85386
20	1676	57.7	732	16	AAV85385
21	1665	57.3	532	21	AAV58162
22	1665	57.3	532	23	AAU11407
23	1665	57.3	588	21	AAV58161
24	1665	57.3	588	23	AAU11406
25	1665	57.3	724	21	AAV58160
26	1665	57.3	724	23	AAU11405
27	491.5	16.9	756	21	AAV71231
28	490	16.9	554	16	AAV08987
29	490	16.9	554	20	AAV23230
30	490	16.9	781	16	AAV08986
31	490	16.9	781	20	AAV23227
32	476.5	16.4	543	12	AAV13405
33	304.5	10.5	264	12	AAV13407
34	295.5	10.2	370	12	AAV13406
35	234.5	8.1	579	13	AAV29079
36	228.5	7.9	686	5	AAV40068
37	209.5	7.2	579	14	AAV38697
38	205.5	7.1	579	14	AAV38702
39	205.5	7.1	579	17	AAV9721
40	205.5	7.1	598	8	AAV70500
41	200.5	6.9	579	10	AAV94798
42	198	6.8	584	14	AAV30811
43	194	6.7	584	22	AAV04302
44	189	6.5	584	22	AAV50109
45	188	6.5	584	23	AAV76051

ALIGNMENTS

RESULT 1
AAV71169
ID AAV71169 standard; Protein; 534 AA.
XX
AC AAV71169;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 capsid protein VP3.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KM cap protein; recombinant viral vector; gene delivery; gene therapy;
KM vaccine; transgene; VP3.
XX
OS Adeno associated virus serotype 1.
XX
PN W0200028061.A2.
XX
PD 18-MAY-2000.
XX
PF 02-NOV-1999; 99MO-US25694.
XX
PR 05-NOV-1998; 98US-0107114.
XX
PI (UNIPROT) UNIV PENNSYLVANIA.
XX
PI Wilson JM, Xiao W;
XX
DR WPI; 2000-376571/32.
XX
DR N-PSDB; AAD00772, AAD00779.
XX
PT Novel adeno-associated virus serotype 1 polynucleotide useful for
preparation of medicament for delivery of a transgene to a host -

XX Claim 7; Page 99-101; 108bp; English.

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
CC which is characterized by two inverted terminal repeats (ITR) and open
CC reading frames for rep and capsid (cap) proteins. The rep reading frame
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
CC coding regions, are useful in production of recombinant viral vectors
CC for gene delivery. These vectors can be used as gene therapy
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC not induce the formation of neutralising antibodies specific to any
CC serotype of AAV hence is useful for transforming host cells, and in the
CC preparation of a medicament for the delivery of transgene to a host.
CC The present sequence is an AAV-1 cap protein VP3 which is
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 534 AA;

Query Match 100.0%; Score 2906; DB 21; Length 534;
Best Local Similarity 100.0%; Pred. No. 8,4e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNHLYKOISS 60
DB 1 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNHLYKOISS 60
QY 61 ASTGASNDNHFGYSTPWFDFPDRFCHSPRDMQRLINNMGPRKRLNFKLFINIQVK 120
DB 61 ASTGASNDNHFGYSTPWFDFPDRFCHSPRDMQRLINNMGPRKRLNFKLFINIQVK 120
QY 121 EVTTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 180
DB 121 EVTTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 180
QY 181 NSGOAVGRSSFCLEFPFSQMLRTGNFTFSYTEEVPFHSVAHSOSLDRLNPLIDQY 240
DB 181 NSGOAVGRSSFCLEFPFSQMLRTGNFTFSYTEEVPFHSVAHSOSLDRLNPLIDQY 240
QY 241 LYYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSFT 300
DB 241 LYYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSFT 300
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNWITD 360
DB 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNWITD 360
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAMGALPGMWODRDVYLQGPIMAKI 420
DB 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAMGALPGMWODRDVYLQGPIMAKI 420
QY 421 PHTDGHFHPSPLMGFGGLKNPPQILLKNTVPANPAPESATKFASTIITQYSTGOVSVE 480
DB 421 PHTDGHFHPSPLMGFGGLKNPPQILLKNTVPANPAPESATKFASTIITQYSTGOVSVE 480
QY 481 IEMELOKENSRRNPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLRPL 534
DB 481 IEMELOKENSRRNPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLRPL 534

RESULT 2

AA71168 standard; Protein; 599 AA.

AA71168;

08-SEP-2000 (first entry)

Adeno-associated virus serotype 1 capsid protein VP2.

Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
cap protein; recombinant viral vector; gene delivery; gene therapy;

KM Vaccine; transgene; VP2.

OS Adeno associated virus serotype 1.

XX W0200028061-A2.

PD 18-MAY-2000.

XX 02-NOV-1999; 99NO-US25694.

PR 05-NOV-1998; 98US-0107114.

PA (UNIV PENNSYLVANIA.

PI Wilson JM, Xiao W;

XX WPI; 2000-376571/32.

DR N-PSDB; AAD00772, AAD00778.

PT Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host

PS Claim 7; Page 93-95; 108bp; English.

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
CC which is characterised by two inverted terminal repeats (ITR) and open
CC reading frames for rep and capsid (cap) proteins. The rep reading frame
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
CC coding regions, are useful in production of recombinant viral vectors
CC for gene delivery. These vectors can be used as gene therapy
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC not induce the formation of neutralising antibodies specific to any
CC serotype of AAV hence is useful for transforming host cells, and in the
CC preparation of a medicament for the delivery of transgene to a host.
CC The present sequence is an AAV-1 cap protein VP2 which is
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 599 AA;

Query Match 100.0%; Score 2906; DB 21; Length 599;
Best Local Similarity 100.0%; Pred. No. 9,9e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNHLYKOISS 60
DB 66 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNHLYKOISS 125
QY 61 ASTGASNDNHFGYSTPWFDFPDRFCHSPRDMQRLINNMGPRKRLNFKLFINIQVK 120
DB 126 ASTGASNDNHFGYSTPWFDFPDRFCHSPRDMQRLINNMGPRKRLNFKLFINIQVK 185
QY 121 EVTTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 180
DB 121 EVTTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 245
QY 186 ETTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 245
DB 186 ETTNDGVTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPADVFMIPQGYITLN 245
QY 181 NSGOAVGRSSFCLEFPFSQMLRTGNFTFSYTEEVPFHSVAHSOSLDRLNPLIDQY 240
DB 246 NSGOAVGRSSFCLEFPFSQMLRTGNFTFSYTEEVPFHSVAHSOSLDRLNPLIDQY 305
QY 241 LYYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSFT 300
DB 306 LYYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSFT 365
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNWITD 360
DB 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNWITD 425
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAMGALPGMWODRDVYLQGPIMAKI 420
DB 426 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHAMGALPGMWODRDVYLQGPIMAKI 485

QY 421 PHTDGHFHPSPMLMGFGGLKKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQSVSE 480
 DB 486 PHTDGHFHPSPMLMGFGGLKKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQSVSE 545
 QY 481 IEMELQKENSCKRMNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLTRPL 534
 DB 546 IEMELQKENSCKRMNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLTRPL 599

RESULT 3
 AAY71167
 ID AAY71167 standard; Protein; 736 AA.
 XX
 AC AAY71167;
 XX
 DT 08-SEP-2000 (first entry)
 XX
 DE Adeno-associated virus serotype 1 capsid protein VP1.
 XX
 KM Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
 KM cap protein; recombinant viral vector; gene delivery; gene therapy;
 KM vaccine; transgene; VPI.
 XX
 OS Adeno associated virus serotype 1.
 XX
 PN W0200028061-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 02-NOV-1999; 99MO-US25694.
 XX
 PR 05-NOV-1998; 98US-0107114.
 XX
 PA (TYPE-) UNIV PENNSYLVANIA.
 XX
 PI Wilson JM, Xiao W;
 XX
 DR WPI; 2000-376571/32.
 DR N-FSDB; AAD00772, AAD00777.
 XX
 PT Novel adeno-associated virus serotype 1 polynucleotide useful for
 preparation of medicament for delivery of a transgene to a host
 XX
 PS Claim 7; Page 87-90; 108pp; English.
 XX
 CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
 CC which is characterised by two inverted terminal repeats (ITR) and open
 CC reading frames for rep and capsid (cap) proteins. The rep reading frame
 CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
 CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
 CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
 CC coding regions, are useful in production of recombinant viral vectors
 CC for gene delivery. These vectors can be used as gene therapy
 CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
 CC not induce the formation of neutralising antibodies specific to any
 CC serotype of AAV hence is useful for transforming host cells, and in the
 CC preparation of a medicament for the delivery of transgene to a host.
 CC The present sequence is an AAV-1 cap protein VPI which is
 CC useful in the production of recombinant viral vector for gene delivery.
 CC
 SQ Sequence 736 AA;
 Query Match 100.0%; Score 2906; DB 21; Length 736;
 Best Local Similarity 100.0%; Pred. No. 1.3e-236;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGAGPAMADNNEGADGVGNASGNMHCDSITWLGDRIVITSTRWALPTYNHLYKQISS 60
 DB 203 MASGGAGPAMADNNEGADGVGNASGNMHCDSITWLGDRIVITSTRWALPTYNHLYKQISS 262
 QY 61 ASTGASNDNFYFGYSTWGYFDENRFCHFSPRDMQRLINNMGFRPKRLNFKLENIQVK 120
 DB 263 ASTGASNDNFYFGYSTWGYFDENRFCHFSPRDMQRLINNMGFRPKRLNFKLENIQVK 322

QY 121 EYTTNDGVTTIANNLSTVOVPSDSEYOLPYVIGASNOGCLPPFPADVEMIPOYGYLTIN 180
 DB 323 EYTTNDGVTTIANNLSTVOVPSDSEYOLPYVIGASNOGCLPPFPADVEMIPOYGYLTIN 382
 QY 181 NSQAVGRSSFYCLEYFSPQMLRTGNFTFSYFEEVPHSSYAHQSQSLDRMLNPLIDQY 240
 DB 383 NSQAVGRSSFYCLEYFSPQMLRTGNFTFSYFEEVPHSSYAHQSQSLDRMLNPLIDQY 442
 QY 241 LYIANTONQSGAOKNDLIFSRGSPAGMSVOPKXMLPGPCTRQOQVSKTKTDNNNSFT 300
 DB 443 LYIANTONQSGAOKNDLIFSRGSPAGMSVOPKXMLPGPCYRQOQVSKTKTDNNNSFT 502
 QY 301 WTGASKYXNLNGRESIINPGTAMASHDDDKFPPMSGWNIPGKESAGASTALDNTMTD 360
 DB 503 WTGASKYXNLNGRESIINPGTAMASHDDDKFPPMSGWNIPGKESAGASTALDNTMTD 562
 QY 361 EBEIKATNPVATERFGTAVNPFQSSSTDPAIGDVHAMGALPGMWODRDVYLQGPIMAKI 420
 DB 563 EBEIKATNPVATERFGTAVNPFQSSSTDPAIGDVHAMGALPGMWODRDVYLQGPIMAKI 622
 QY 421 PHTDGHFHPSPMLMGFGGLKKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQSVSE 480
 DB 623 PHTDGHFHPSPMLMGFGGLKKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQSVSE 682

QY 481 IEMELQKENSCKRMNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLTRPL 534
 DB 683 IEMELQKENSCKRMNPEVOYTSNYSKASANDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 4
 AAB59847
 ID AAB59847 standard; Protein; 736 AA.
 XX
 AC AAB59847;
 XX
 DT 28-MAR-2001 (first entry)
 XX
 DE AAV6 capsid protein VP1.
 XX
 KM AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 KM atherosclerosis; sickle cell anaemia; thalassaemia;
 KM blood clotting disorder; diabetes; capsid protein VP1.
 XX
 OS Adeno associated virus.
 XX
 PN US6156303-A.
 XX
 PD 05-DEC-2000.
 XX
 PF 11-JUN-1997; 97US-0873168.
 XX
 PR 11-JUN-1997; 97US-0873168.
 XX
 PA (UNIV) UNIV WASHINGTON.
 XX
 PI Russell DW, Rutledge EA;
 XX
 DR WPI; 2001-060164/07.
 XX
 PT Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalassaemia and diabetes
 XX
 PS Disclosure; Fig 2; 50pp; English.
 XX
 CC The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome

CC can be modified to express a capsid protein of an AAV serotype that
transduces the selected host cell.

XX Sequence 736 AA;

Query Match 99.2%; Score 2884; DB 22; Length 736;
Best Local Similarity 99.1%; Pred. No. 9,6e-235;
Matches 529; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADGVGNASGNWCHDSITWLGDRVITTSRTTWTALPTYNNHLYKQISS 60
DB 203 MASGGAPMADNNEGADGVGNASGNWCHDSITWLGDRVITTSRTTWTALPTYNNHLYKQISS 262
QY 61 ASTGASDNHNYFGYSTPMWGFDFENRPHCHSPRDMORLINNNGFPRKRLNFKLFNIQVK 120
DB 263 ASTGASDNHNYFGYSTPMWGFDFENRPHCHSPRDMORLINNNGFPRKRLNFKLFNIQVK 322
QY 121 EVTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 180
DB 323 EVTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 382
QY 181 NGSOAVGRSSFFCYCLEYFPPSQMLRTGNNTFSEYFEEVFPFHSYAHQSGLDRLANPLIDQY 240
DB 383 NGSOAVGRSSFFCYCLEYFPPSQMLRTGNNTFSEYFEEVFPFHSYAHQSGLDRLANPLIDQY 442
QY 241 LYLINRTQNGSAGNOKDILFSGSPAGMSVOPKMWLPGPCYRQORVSKTKTDNNNSNF 300
DB 443 LYLINRTQNGSAGNOKDILFSGSPAGMSVOPKMWLPGPCYRQORVSKTKTDNNNSNF 502
QY 301 WTGASKTNLNGRESIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDNVMT 360
DB 503 WTGASKTNLNGRESIINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDNVMT 562
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAFGDVHAMGALPGMWQDRDVLQGPIMAKI 420
DB 563 EEEIKATNPVATERFGTVAVNFQSSSTDPAFGDVHAMGALPGMWQDRDVLQGPIMAKI 622
QY 421 PHTDGHFHPSPIMGFGGLKNPPQILIKNTVPANPAPAEFSATKPAFTIYOSTGQVSV 480
DB 623 PHTDGHFHPSPIMGFGGLKNPPQILIKNTVPANPAPAEFSATKPAFTIYOSTGQVSV 682
QY 481 EEMELQKENSKRMPNEVOYTSNYAKSANVDFVNNGLYTEPRDIGRILTRPL 534
DB 683 EEMELQKENSKRMPNEVOYTSNYAKSANVDFVNNGLYTEPRDIGRILTRPL 736

RESULT 5
ID AAB59846 standard; Protein; 736 AA.

XX AAB59846;
XX 28-MAR-2001 (first entry)
XX AAV3B capsid protein VP1.
XX AAV3B: gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX atherosclerosis; sickle cell anaemia; thalassemia;
XX blood clotting disorder; diabetes; capsid protein VP1.
XX Adeno associated virus.
XX OS
XX PN US6156303-A.
XX PD 05-DEC-2000.
XX PF 11-JUN-1997; 97US-0873168.
XX PR 11-JUN-1997; 97US-0873168.
XX PA (UNITW) UNIV WASHINGTON.
XX PI Russell DW, Rutledge EA;

XX WPI; 2001-060164/07.

XX Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX Disclosure; Fig 2; 50pp; English.

XX The present invention relates to adeno-associated virus serotypes. The
XX present sequence is capsid protein VP1 of one such serotype (AAV3B).
XX AAV3B can be used to construct AAV viral vectors for use in gene therapy
XX for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.

XX Sequence 736 AA;

Query Match 87.1%; Score 2530; DB 22; Length 736;
Best Local Similarity 86.0%; Pred. No. 7,8e-205;
Matches 460; Conservative 27; Mismatches 46; Indels 2; Gaps 2;

QY 1 MASGGAPMADNNEGADGVGNASGNWCHDSITWLGDRVITTSRTTWTALPTYNNHLYKQISS 60
DB 203 MASGGAPMADNNEGADGVGNASGNWCHDSITWLGDRVITTSRTTWTALPTYNNHLYKQISS 262
QY 61 ASTGASDNHNYFGYSTPMWGFDFENRPHCHSPRDMORLINNNGFPRKRLNFKLFNIQVK 120
DB 263 AS-GASDNHNYFGYSTPMWGFDFENRPHCHSPRDMORLINNNGFPRKRLNFKLFNIQVK 321
QY 121 EVTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 180
DB 322 EVTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 381
QY 181 NGSOAVGRSSFFCYCLEYFPPSQMLRTGNNTFSEYFEEVFPFHSYAHQSGLDRLANPLIDQY 240
DB 382 NGSOAVGRSSFFCYCLEYFPPSQMLRTGNNTFSEYFEEVFPFHSYAHQSGLDRLANPLIDQY 441
QY 421 PHTDGHFHPSPIMGFGGLKNPPQILIKNTVPANPAPAEFSATKPAFTIYOSTGQVSV 480
DB 422 PHTDGHFHPSPIMGFGGLKNPPQILIKNTVPANPAPAEFSATKPAFTIYOSTGQVSV 681
QY 480 EEMELQKENSKRMPNEVOYTSNYAKSANVDFVNNGLYTEPRDIGRILTRPL 534
DB 682 EEMELQKENSKRMPNEVOYTSNYAKSANVDFVNNGLYTEPRDIGRILTRPL 736

RESULT 6

ID AAB59845 standard; Protein; 736 AA.

XX AAB59845;
XX 28-MAR-2001 (first entry)
XX AAV3A capsid protein VP1.
XX AAV3A: gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

KW atherosclerosis; sickle cell anaemia; thalassaemia;
 KW blood clotting disorder; diabetes; capsid protein VP1.
 XX Adeno associated virus.
 OS US6156303-A.
 XX
 XX
 XX 05-DEC-2000.
 XX
 XX 11-JUN-1997; 97US-0873168.
 XX
 XX 11-JUN-1997; 97US-0873168.
 XX
 XX (UNIW) UNIV WASHINGTON.
 XX
 XX Russell DW, Rutledge EA;
 DR WPI; 2001-060164/07.
 XX
 XX Adeno-associated virus serotype 6 and viral vector derived from it for
 PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 PT syndrome, sickle cell anemia, thalassaemia and diabetes
 XX
 XX Disclosure; Fig 2; 50pp; English.
 XX
 XX The present invention relates to adeno-associated virus serotypes. The
 CC present sequence is capsid protein VP1 of one such serotype (AAV3A).
 CC AAV3A can be used to construct AAV viral vectors for use in gene therapy
 CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.
 XX
 XX Sequence 736 AA;
 SQ
 Query Match 87.0%; Score 2528; DB 22; Length 736;
 Best Local Similarity 85.8%; Pred. No. 1.2e-204;
 Matches 445; Conservative 28; Mismatches 46; Indels 2; Gaps 2;
 QY 1 MASGGGAPMADNNEGADGVNAGSNMHCDSITWLDGRTVITSTRWALPTYNHLYKQISS 60
 DB 203 MASGGGAPMADNNEGADGVNAGSNMHCDSITWLDGRTVITSTRWALPTYNHLYKQISS 262
 QY 61 ASIGASNDNHYFGYSTWGFDFRRFCHSPRDMQGLINNNGFRKXLFNFKLPIQVK 120
 DB 263 QS-GASNDNHYFGYSTWGFDFRRFCHSPRDMQGLINNNGFRKXLFNFKLPIQVK 321
 QY 121 ETTNDGVTIANNLTSTVOVFSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 180
 DB 322 GVTQNDGVTIANNLTSTVOVFSSEYQLPYVLGSAHQGLPPPPADVFMIPOYGYLTIN 381
 QY 181 NSGQAVGRSSFFCYLEYPSPQMLRTGNNFTSYFEEVFFHSSYAHSGSLDRMLNPLIDY 240
 DB 382 NSGQAVGRSSFFCYLEYPSPQMLRTGNNFTSYFEEVFFHSSYAHSGSLDRMLNPLIDY 441
 QY 241 LYTIANRQ-NQSSAQKMLLFSRGSAGMSVDPKXMLPBPCTROQVSKTKIDNNNSNF 299
 DB 442 LYTIANRQ-NQSSAQKMLLFSRGSAGMSVDPKXMLPBPCTROQVSKTKIDNNNSNF 501
 QY 300 TWTGASRYNINRGRESIINPGTAMASHKDDKFFPMGVMIFGKESAGSANTLADNMIT 359
 DB 502 PWTAASTKYLNGRDSLVPNPGPMAHSHDDDEKFFPMHGNLIFKRGKTTASNAELDNVIT 561
 QY 360 DEEIRKATNPVATERFGTVAVNFQSSSTDPATGDAHAMGALPGVMWQDRDVTYLGPIWAK 419
 DB 562 DEEIRKATNPVATERFGTVAVNFQSSSTDPATGDAHAMGALPGVMWQDRDVTYLGPIWAK 621
 QY 420 IPRTHGHHFHSPLMGSGGLKNPPQILLKNTPPANPAFSAFKPSFTTQYSTGVSV 479
 DB 622 IPRTHGHHFHSPLMGSGGLKNPPQILLKNTPPANPAFSAFKPSFTTQYSTGVSV 681

QY 480 EIEMLQKNSKRNPEVOYTSNYAKSANVDFVNDNGHYTEBRPIGRTYLRPL 534
 DB 682 EIEMLQKNSKRNPEVOYTSNYAKSANVDFVNDNGHYTEBRPIGRTYLRPL 736
 RESULT 7
 ID AAG65794
 XX AAG65794 standard; Protein; 533 AA.
 AC AAG65794;
 XX
 XX 11-FEB-2002 (first entry)
 XX
 XX Adeno-associated virus 2 (AAV-2) major coat protein VP3.
 DE
 XX
 KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
 KW inverted terminal repeat; nonretroviral; neuroprotective; antineoplastic; IRK;
 KW antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
 KW major coat protein; AAV-2; VP3.
 XX
 XX Adeno-associated virus 2.
 OS
 XX
 XX WO20016888-A2.
 XX
 XX 20-SEP-2001.
 XX
 XX 13-MAR-2001; 2001MO-US07927.
 XX
 XX 14-MAR-2000; 2000US-189110P.
 XX
 XX (NEUR-) NEUROLOGIX INC.
 XX
 XX Xiao W, During MJ;
 PI WPI; 2001-596912/67.
 XX
 XX N-PSDB; AAI66974.
 DR
 XX
 XX Recombinant viral vector useful in improving gene therapy in a subject,
 PT and for increasing efficiency of entry into a cell, comprises a desired
 PT chimeric capsid having one non-native amino acid sequence and a desired
 PT transgene
 XX
 XX Disclosure; Page 51; 53pp; English.
 XX
 XX The invention provides a recombinant viral vector (RVV) comprising a
 CC chimeric capsid (I) having at least one non-native amino acid sequence,
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),
 CC or their combination, and a transgene flanked 5' and 3' by inverted
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their
 CC combination. The RVV is useful for improving gene therapy in a subject
 CC with a disorder, and for increasing the efficiency of entry into a cell,
 CC which involves producing (I) encapsulating a viral vector, and contacting
 CC a cell with RVV having (I) such that (I) binds to an attachment site on
 CC the cell surface and permits the vector to enter the cell efficiently.
 CC A pharmaceutical composition comprising RVV with (I) containing a
 CC transgene sequence associated with a disease or a disorder such that
 CC expression of the transgene would result in amelioration of the disease
 CC or disorder such as inherited neurological and metabolic diseases e.g.
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,
 CC retinoblastoma and various types of neoplastic cells which include
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,
 CC sarcomas, leukemias and lymphoma. The present sequence represents the
 CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences
 CC are used in the construction of a chimeric vector.
 XX
 SO Sequence 533 AA;
 Query Match 85.4%; Score 2481.5; DB 22; Length 533;
 Best Local Similarity 83.3%; Pred. No. 6.2e-201;
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNNHLYKQISS 60
PT 1 MATSGGAPMADNNEGADGVGNASGNMHCDSITWMDRVTITSTRWALPTYNNHLYKQISS 60
PT 1 MATSGGAPMADNNEGADGVGNASGNMHCDSITWMDRVTITSTRWALPTYNNHLYKQISS 60
QY 61 ASTGASNDNHYFGYSTPMGYDFNRFCHFSPRDMORLNNNNKPRKRLNFKLEFNIOVK 120
DB 61 OS-GASNDNHYFGYSTPMGYDFNRFCHFSPRDMORLNNNNKPRKRLNFKLEFNIOVK 119
QY 121 EVTTNDGVTITANNLTSTVOVFSSEYOLPYVLSAHQGLPPEPPADVFEMIPQYGLTLN 180
DB 121 EVTTNDGVTITANNLTSTVOVFSSEYOLPYVLSAHQGLPPEPPADVFEMIPQYGLTLN 179
QY 181 NSGQAVGRSSFYCLEYFPSSQMLRTGNNTFSEYFEEVPHSSYAHQSOLDRMLNPLIDQY 240
DB 180 NSGQAVGRSSFYCLEYFPSSQMLRTGNNTFSEYFEEVPHSSYAHQSOLDRMLNPLIDQY 239
QY 241 LYYLNTQNGSGAQNKDILFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300
DB 240 LYYLNTQNGSGAQNKDILFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 299
QY 301 WTGASKTNLNGRESITINPGTAMASHKODEKFPKMSGMTFGKESAGASNTALDNMTD 360
DB 300 WTGASKTNLNGRESITINPGTAMASHKODEKFPKMSGMTFGKESAGASNTALDNMTD 359
QY 361 EEEIKATNPVATERFGTAVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLQGPWAKI 420
DB 360 EEEIKATNPVATERFGTAVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLQGPWAKI 419
QY 421 PHTDGHFHPSPLMGFGGLKNPPOILLIKNTVPANPAPESATKFASTIQYSTGVSV 480
DB 420 PHTDGHFHPSPLMGFGGLKNPPOILLIKNTVPANPAPESATKFASTIQYSTGVSV 479
QY 481 IEMELQKENSKRNNPEVOYTSNAYKASANDFTVNNGLYTERPPICGRYLRPL 534
DB 480 IEMELQKENSKRNNPEVOYTSNAYKASANDFTVNNGLYTERPPICGRYLRPL 533

RESULT 8
AAG65793
ID AAG65793 standard; Protein; 598 AA.
AC AAG65793;
XX
DT 11-FEB-2002 (first entry)
XX
DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.
XX
KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
KM inverted terminal repeat; nonretroviral; antineoplastic; ITR;
KM antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
KM major coat protein; AAV-2; VP2.
XX
OS Adeno-associated virus 2.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "encoded by ACG"
XX
XX WO200168888-A2.
XX
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001WO-US07927.
XX
PR 14-MAR-2000; 2000US-189110P.
XX
PA (NEUR-) NEUROLOGIX INC.
XX
PI Xiao W, During MJ;
XX WPI; 2001-596912/67.
DR N-PSDB; AAI66974.

XX
PT Recombinant viral vector useful in improving gene therapy in a subject,
PT and for increasing efficiency of entry into a cell, comprises a
PT chimeric capsid having one non-native amino acid sequence and a desired
PT transgene
PS Disclosure; Page 51; 53pb; English.
XX
CC The invention provides a recombinant viral vector (RVV) comprising a
CC chimeric capsid (I) having at least one non-native amino acid sequence,
CC derived from a capsid protein domain of parvovirus (II), a virus (III),
CC or their combination, and a transgene flanked 5' and 3' by inverted
CC terminal repeat (ITR) sequences, derived from (II), (III), or their
CC combination. The RVV is useful for improving gene therapy in a subject
CC with a disorder, and for increasing the efficiency of entry into a cell,
CC which involves producing (I) encapsulating a viral vector, and contacting
CC a cell with RVV having (I) such that (I) binds to an attachment site on
CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (I) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Leach-Nyhan syndrome, amyloid polynuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include
CC tumours especially central nervous system tumours, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences
CC are used in the construction of a chimeric vector.
SQ Sequence 598 AA;
Query Match 85.4%; Score 2481.5; DB 22; Length 598;
Best Local Similarity 83.3%; Pred. No. 7.3e-201;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;
QY 1 MASGGGAPMADNNEGADGVGNASGNMHCDSITWLGDRVITTSRTWALPTYNNHLYKQISS 60
DB 66 MATSGGAPMADNNEGADGVGNASGNMHCDSITWMDRVTITSTRWALPTYNNHLYKQISS 125
QY 61 ASTGASNDNHYFGYSTPMGYDFNRFCHFSPRDMORLNNNNKPRKRLNFKLEFNIOVK 120
DB 126 OS-GASNDNHYFGYSTPMGYDFNRFCHFSPRDMORLNNNNKPRKRLNFKLEFNIOVK 184
QY 121 EVTTNDGVTITANNLTSTVOVFSSEYOLPYVLSAHQGLPPEPPADVFEMIPQYGLTLN 180
DB 185 EVTTNDGVTITANNLTSTVOVFSSEYOLPYVLSAHQGLPPEPPADVFEMIPQYGLTLN 244
QY 181 NSGQAVGRSSFYCLEYFPSSQMLRTGNNTFSEYFEEVPHSSYAHQSOLDRMLNPLIDQY 240
DB 245 NSGQAVGRSSFYCLEYFPSSQMLRTGNNTFSEYFEEVPHSSYAHQSOLDRMLNPLIDQY 304
QY 241 LYYLNTQNGSGAQNKDILFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300
DB 305 LYYLNTQNGSGAQNKDILFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 364
QY 301 WTGASKTNLNGRESITINPGTAMASHKODEKFPKMSGMTFGKESAGASNTALDNMTD 360
DB 365 WTGASKTNLNGRESITINPGTAMASHKODEKFPKMSGMTFGKESAGASNTALDNMTD 424
QY 421 PHTDGHFHPSPLMGFGGLKNPPOILLIKNTVPANPAPESATKFASTIQYSTGVSV 480
DB 485 PHTDGHFHPSPLMGFGGLKNPPOILLIKNTVPANPAPESATKFASTIQYSTGVSV 544
QY 481 IEMELQKENSKRNNPEVOYTSNAYKASANDFTVNNGLYTERPPICGRYLRPL 534
DB 545 IEMELQKENSKRNNPEVOYTSNAYKASANDFTVNNGLYTERPPICGRYLRPL 598

RESULT 9
AAM51509
ID AAM51509 standard; protein; 598 AA.
XX
AC AAM51509;
XX
DT 02-JAN-2002 (first entry)
XX
DE Adeno-associated virus VP2 capsid protein.
XX
KM Adeno-associated virus; AAV; VP2; capsid; virus-like particle;
KM nuclear localisation signal; VP3.
XX
OS Adeno associated virus.
XX
PN JP2001169777-A.
XX
PD 26-JUN-2001.
XX
PF 30-JUL-1999; 99JP-0249140.
XX
PR 30-JUL-1999; 99JP-0249140.
XX
PA (HAND/) HANDA H.
XX
DR WPI; 2001-559854/68.
XX
PT New virus-like particles from VP3 capsid protein of adeno-associated
PT virus, comprise a peptide containing a nucleus-shifting signal
PT connected to its N-terminal -
XX
PS Claim 1; Page 14-16; 33pp; Japanese.
XX
CC The present sequence is provided in a specification relating to a
CC virus-like particle-forming protein, and to a peptide containing a
CC nuclear-shifting signal at its N-terminus that can form virus-like
CC particles by shifting to the nucleus of the animal cell in which it
CC is expressed. The method is used for forming virus-like particles
CC from the VP3 capsid protein of adeno-associated virus (AAV).
XX
SQ Sequence 598 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 598;
Best Local Similarity 83.3%; Pred. No. 7.3e-201;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVNAGNMHCDSTWLGDRVITTSRTWALPTYNHLYKOISS 60
DB 66 MATGSGAPMADNNEGADGVNAGNMHCDSTWMDRITTSRTWALPTYNHLYKOISS 125
QY 61 ASFGASNDNHFGYSTFWGYFDENRFCHFSRPRDQRLINNNMGFRPRKLNFKLNIOVK 120
DB 126 QS-GASNDNHFGYSTFWGYFDENRFCHFSRPRDQRLINNNMGFRPRKLNFKLNIOVK 184
QY 121 EVTTTNGCVTTIANNLTSTVOVFSSEYQLPYVLSAIOGCLPPRPADVNIPOGYITLN 180
DB 185 EVTQNGCVTTIANNLTSTVOVFSSEYQLPYVLSAIOGCLPPRPADVNIPOGYITLN 244
QY 181 NSGOAVRSRFFCYCLEYFPSPQMLRTGNFTFSYTFEEVPFHSYAHOSQLRLNMPLIQY 240
DB 245 NSGOAVRSRFFCYCLEYFPSPQMLRTGNFTFSYTFEEVPFHSYAHOSQLRLNMPLIQY 304
QY 241 LVIYIANTQNSGSAONKDLFSGKSPAGMSVOPKRWLPGBCYRQQRVSKTKTNNNSFT 300
DB 305 LVIYIANTQNSGSAONKDLFSGKSPAGMSVOPKRWLPGBCYRQQRVSKTKTNNNSFT 364
QY 301 WTGASKTNLNGRRESIINPGTAMASHKDDKDFPMGVMIFGKESASASTALDNTWITD 360
DB 365 WTGATKTHLNGRDSLVNPGPMASHKDDKDFPMGVMIFGKESASASTALDNTWITD 424
QY 361 EEBIKATNPVATERFGTVAVNFQSSSTDPATGDVHAGALPGMWQDRDVLQGPWAKI 420

DB 425 EEBIKATNPVATERFGTVAVNFQSSSTDPATGDVHAGALPGMWQDRDVLQGPWAKI 484
QY 421 PHTDGHFHPSPMLMGSGGLKRNPPQIILIKTTPVANPAPAESAKFKSFITQYSTGVSE 480
DB 485 PHTDGHFHPSPMLMGSGGLKRNPPQIILIKTTPVANPAPAESAKFKSFITQYSTGVSE 544
QY 481 IEMWLOKENSKRWNPEVOYTSNYSKANSVDFTDNNGLYTERPPIGRTYTRPL 534
DB 545 IEMWLOKENSKRWNPEVOYTSNYSKANSVDFTDNNGLYTERPPIGRTYTRPL 598

RESULT 10
AAG65792
ID AAG65792 standard; Protein; 735 AA.
XX
AC AAG65792;
XX
DT 11-FEB-2002 (first entry)
XX
DE Adeno-associated virus 2 (AAV-2) major coat protein VP1.
XX
KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;
KM inverted terminal repeat; nonretroviral; neuroprotective; anti-neurotic; ITR;
KM antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV;
KM major coat protein; AAV-2; VP1.
XX
OS Adeno-associated virus 2.
XX
PN WO200168888-A2.
XX
PD 20-SEP-2001.
XX
PF 13-MAR-2001; 2001MO-US07927.
XX
PR 14-MAR-2000; 2000US-189110P.
XX
PA (NEUR-) NEUROLOGIX INC.
XX
PI Xiao W, During MJ;
XX
DR WPI; 2001-5596912/67.
DR N-PSDB; AAI66974.
XX
PT Recombinant viral vector useful in improving gene therapy in a subject,
PT and for increasing efficiency of entry into a cell, comprises a
PT chimeric capsid having one non-native amino acid sequence and a desired
PT transgene -
XX
PS Disclosure; Page 50; 53pp; English.

XX The invention provides a recombinant viral vector (RVV) comprising a
CC chimeric capsid (I) having at least one non-native amino acid sequence,
CC derived from a capsid protein domain of parvovirus (II), a virus (III),
CC or their combination, and a transgene flanked 5' and 3' by inverted
CC terminal repeat (ITR) sequences, derived from (II), (III), or their
CC combination. The RVV is useful for improving gene therapy in a subject
CC with a disorder, and for increasing the efficiency of entry into a cell,
CC which involves producing (I) encapsulating a viral vector, and contacting
CC a cell with RVV having (I) such that (I) binds to an attachment site on
CC the cell surface and permits the vector to enter the cell efficiently.
CC A pharmaceutical composition comprising RVV with (I) containing a
CC transgene sequence associated with a disease or a disorder such that
CC expression of the transgene would result in amelioration of the disease
CC or disorder such as inherited neurological and metabolic diseases e.g.
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polyneuropathy,
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
CC fibrosis, diabetes, diseases associated with hormone deficiencies,
CC retinoblastoma and various types of neoplastic cells which include
CC tumours especially central nervous system tumours, neoplasms, carcinomas,
CC sarcomas, leukemias and lymphoma. The present sequence represents the
CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences

CC are used in the construction of a chimeric vector.

SQ Sequence 735 AA;

Query Match	85.4%;	Score 2481.5;	DB 22;	Length 735;
Best Local Similarity	83.3%;	Pred. No. 9.8e-201;		
Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;

Qy	1	MASGGAPMADNNNEGADGVAGNAGNMWCDSTWLGDRVITTSRTTALPTYNHLYKQISS	60
Db	203	MATSGAPMADNNNEGADGVAGNAGNMWCDSTWLGDRVITTSRTTALPTYNHLYKQISS	262
Qy	61	ASTGASNDNHVFGYSTPWGYDFDNFCHCFSPRDMQRLINNNGFRPKRLNFKLENIOVK	120
Db	263	QS-GASNDNHVFGYSTPWGYDFDNFCHCFSPRDMQRLINNNGFRPKRLNFKLENIOVK	321
Qy	121	EVTINDGYTTIANNLTSTVOYFSDSEYQLPYLGSAGHQCCLPFPADVFPMIPQYGLTLN	180
Db	322	EVTQNDGTTTIANNLTSTVOYFTDSEYQLPYLGSAGHQCCLPFPADVFPMIPQYGLTLN	381
Qy	181	NGSOAVGRSSFFCYCLEYFPSSOMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY	240
Db	382	NGSOAVGRSSFFCYCLEYFPSSOMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRMLNPLIDQY	441
Qy	241	LYLINRTNOGSGSAQNKDLLFSRGSFAGMSVQPKMWLPQPCYRQQRVSKTKTDNNNSNFT	300
Db	442	LYLSLRTNTPSGTQTTSRLQFSQAGSIDRDSRAMLPGPCYRQQRVSKTSADNNNSYS	501
Qy	301	WTGASKTYLNGRESLINPGTMAASHKDDDEKFFPMSSGVNIPCKEKASNTLMDVMITD	360
Db	502	WTGAKRYLNGSDSLVNPQPAASHKDDDEKFFPQSGVLIFFKQSGSEKTNVDIEKVMITD	561
Qy	361	EEELKATNPVATEREGTVAVNFPOSSSTDPATGDVAMGALPGVMQODRDVYLQSGIWKI	420
Db	562	EEELRTNPVATEREGTGVSYSTNLQRBGRQAATADVDTQVGLPGVMQODRDVYLQSGIWKI	621
Qy	421	PHTDGHFSPPLMGSGFLKNPPEQLIKNTVPANPAPAEFSATYKASFTQYSTQGVSE	480
Db	622	PHTDGHFSPPLMGSGFLKHPPEQLIKNTVPANPSTTFSSAKFASFTQYSTQGVSE	681
Qy	481	IEMELQKENSKRNPVEVOYTSNYAKSANDVDFVNDNGLYTERPRTGTRLTREPL	534
Db	682	IEMELQKENSKRNPVEIOYTSNYAKSVANDVFDVNDNGVSEPRPIGTRILTYNKL	735

RESULT 11

AAM51508

ID AAM51508 standard; protein, 735 AA.

XX AAM51508;

XX

XX 02-JAN-2002 (first entry)

XX

XX Adeno-associated virus VP1 capsid protein.

XX

XX Adeno-associated virus; AAV; VP1; capsid; virus-like particle;

XX nuclear localisation signal; VP3.

XX

XX Adeno associated virus.

XX

XX JP2001169777-A.

XX

XX 26-JUN-2001.

XX

XX 30-JUL-1999; 99JP-0249140.

XX

XX 30-JUL-1999; 99JP-0249140.

XX

XX (HAND/) HANDA H.

XX

XX WPI; 2001-599854/68.

XX

XX New virus-like particles from VP3 capsid protein of adeno-associated

PT virus, comprise a peptide containing a nucleus-shifting signal connected to its N-terminal -

PS Disclosure; Page 10-13; 33pp; Japanese

CC The present sequence is provided in a specification relating to a
CC virus-like particle-forming protein, and to a peptide containing a
CC nuclear-shifting signal at its N-terminus that can form virus-like
CC particles by shifting to the nucleus of the animal cell in which it
CC is expressed. The method is used for forming virus-like particles
CC from the VP3 capsid protein of adeno-associated virus (AAV).

SQ Sequence 735 AA;

Query Match	85.4%	Score	2481.5	DB	22	Length	735
Best Local Similarity	83.3%	Pred. No.	9	8e-20			
Matches	445	Conservative	37	Mismatches	51	Indels	1
						Gaps	1

[illegible]

PD 05-DEC-2000.
XX
XX 11-JUN-1997; 97US-0873168.
XX
XX 11-JUN-1997; 97US-0873168.
XX
XX (UNITW) UNIV WASHINGTON.
XX
XX Russell DW, Rutledge EA;
XX WPI; 2001-060164/07.
XX
XX Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes
XX
XX Claim 7, Fig 2, 50pp; English.
XX
XX The present invention relates to adeno-associated virus serotypes. The
XX present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX
XX Sequence 735 AA;
SQ
Query Match 85.4%; Score 2481.5; DB 22; Length 735;
Best Local Similarity 83.3%; Pred. No. 9.8e-201;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;
QY 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLYKOISS 60
DB 203 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLYKOISS 262
QY 61 ASTGASNDNHFGYSTPMGYPDENRPFCHSPRDMOQLINNMGFRKRLNFKLENIOYK 120
DB 263 QS-GASNDNHFGYSTPMGYPDENRPFCHSPRDMOQLINNMGFRKRLNFKLENIOYK 321
QY 121 EYVNDGVTTIANNLSTVOVFSDESEYOLPYVLGSAHOGCLPPPADVPMIPQGYLTIN 180
DB 322 EYVNDGVTTIANNLSTVOVFSDESEYOLPYVLGSAHOGCLPPPADVPMIPQGYLTIN 381
QY 181 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEVPPHSSYAHQSGLDRMLNPLIDY 240
DB 382 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEVPPHSSYAHQSGLDRMLNPLIDY 441
QY 241 LYYLNRITNQGSGAOKNDLFSRGSPPAGMSVOPKMWLPBCYRQORSKTKTDNNNSFT 300
DB 442 LYYLNRITNQGSGAOKNDLFSRGSPPAGMSVOPKMWLPBCYRQORSKTKTDNNNSFT 501
QY 301 WTGASRYNLNGRESIINPGTAMASHKDDDEKFFPMGSMVIFGKESAGASNTALDNVMTD 360
DB 502 WTGATKXHLNGRDSLVNPGPAMASHKDDDEKFFPMGSMVIFGKESAGASNTALDNVMTD 561
QY 361 EEEIKATNPVATERFETVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLOGPYAKI 420
DB 562 EEEIRITNPVATERFETVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLOGPYAKI 621
QY 421 PHTDGHFHPSPILMGFGGLKNPPOILLKNTVPANPAEFSATKFPSTIOYSTGVAVE 480
DB 622 PHTDGHFHPSPILMGFGGLKNPPOILLKNTVPANPAEFSATKFPSTIOYSTGVAVE 681
QY 481 IEMELQENSKRNWPEVOYTSNTAKSANVDFTVNNGLYTEPRPIGTIRYLTREP 534
DB 682 IEMELQENSKRNWPEVOYTSNTAKSANVDFTVNNGLYTEPRPIGTIRYLTREP 935
RESULT 13
AAMS1510

ID AAMS1510 standard; protein; 533 AA.
XX
XX AAMS1510;
AC
XX 02-JAN-2002 (first entry)
DT
XX Adeno-associated virus VP3 capsid protein.
XX
XX Adeno-associated virus VP3 capsid protein.
XX
XX Adeno-associated virus; AAV; capsid; virus-like particle;
XX nuclear localisation signal; VP3.
XX
XX Adeno associated virus.
XX
XX JP2001169777-A.
XX
XX 26-JUN-2001.
PD
XX 30-JUL-1999; 99JP-0249140.
XX
XX 30-JUL-1999; 99JP-0249140.
XX
XX 30-JUL-1999; 99JP-0249140.
XX
XX (HAND/) HANDA H.
XX
XX WPI; 2001-599854/68.
XX
XX New virus-like particles from VP3 capsid protein of adeno-associated
XX virus, comprise a peptide containing a nucleus-shifting signal
XX connected to its N-terminal -
XX
XX Disclosure; Page 17-19; 33pp; Japanese.
XX
XX The present sequence is provided in a specification relating to a
XX virus-like particle-forming protein, and to a peptide containing a
XX nuclear-shifting signal at its N-terminus that can form virus-like
XX particles by shifting to the nucleus of the animal cell in which it
XX is expressed. The method is used for forming virus-like particles
XX from the VP3 capsid protein of adeno-associated virus (AAV).
XX
XX Sequence 533 AA;
SQ
Query Match 85.3%; Score 2477.5; DB 22; Length 533;
Best Local Similarity 83.3%; Pred. No. 1.3e-200;
Matches 445; Conservative 36; Mismatches 52; Indels 1; Gaps 1;
QY 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLYKOISS 60
DB 1 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLYKOISS 60
QY 61 ASTGASNDNHFGYSTPMGYPDENRPFCHSPRDMOQLINNMGFRKRLNFKLENIOYK 120
DB 61 QS-GASNDNHFGYSTPMGYPDENRPFCHSPRDMOQLINNMGFRKRLNFKLENIOYK 119
QY 121 EYVNDGVTTIANNLSTVOVFSDESEYOLPYVLGSAHOGCLPPPADVPMIPQGYLTIN 180
DB 120 EYVNDGVTTIANNLSTVOVFSDESEYOLPYVLGSAHOGCLPPPADVPMIPQGYLTIN 179
QY 181 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEVPPHSSYAHQSGLDRMLNPLIDY 240
DB 180 NSGQAVGRSSFYCLEYFPQMLRTGNNFTFSYFEVPPHSSYAHQSGLDRMLNPLIDY 239
QY 241 LYYLNRITNQGSGAOKNDLFSRGSPPAGMSVOPKMWLPBCYRQORSKTKTDNNNSFT 300
DB 240 LYYLNRITNQGSGAOKNDLFSRGSPPAGMSVOPKMWLPBCYRQORSKTKTDNNNSFT 299
QY 301 WTGASRYNLNGRESIINPGTAMASHKDDDEKFFPMGSMVIFGKESAGASNTALDNVMTD 360
DB 300 WTGATKXHLNGRDSLVNPGPAMASHKDDDEKFFPMGSMVIFGKESAGASNTALDNVMTD 359
QY 361 EEEIKATNPVATERFETVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLOGPYAKI 420
DB 360 EEEIRITNPVATERFETVAVNFQSSSTDPATGDVHAMGALPGMWODRDVYLOGPYAKI 419
QY 421 PHTDGHFHPSPILMGFGGLKNPPOILLKNTVPANPAEFSATKFPSTIOYSTGVAVE 480

Db 420 PHTDGHFHPSPMLGFGFKHPPOILLIKTVPANPSTTFSAKFAFSTTQYSTGVSV 479
QY 481 IEMELOKENSKRNNPEVQYTSNYSKASANDFTDNNGLYTEPRIGTRYLTRL 534
Db 480 IEMELOKENSKRNNPEIQYTSNYSKASVAVDFTVDINGVSEPRIGTRYLTRL 533
RESULT 14
AAB50326
ID AAB50326 standard; protein; 734 AA.
XX AAB50326;
AC AAB50326;
XX 09-MAR-2001 (first entry)
DE Adeno-associated virus capsid protein sequence.
XX Adeno-associated virus; AAV; capsid; virus binding inhibition;
KM competitive inhibitor.
XX Maecadenovirus.
OS Synthetic.
XX WO200073316-A2.
XX 07-DEC-2000.
XX 26-MAY-2000; 2000WO-US14466.
XX 28-MAY-1999; 99US-0321589.
XX (CELL-) CELL GENESYS INC.
PI Patel S, McArthur J;
XX WPI; 2001-061507/07.
XX New polypeptide portion of a virus protein binding to an antibody
PT specific for the virus useful for inhibiting binding of the virus to a
PT cell or for binding host antibody to provide a transient tolerant or
PT non-responsive state
XX Disclosure; Fig 2; 33pp; English.
XX The present sequence is given in a specification relating to
CC polypeptide portions of a virus protein or its derivative, that bind to
CC an antibody specific for the virus or inhibit binding of the virus to a
CC cell. The polypeptides are used in inhibiting the binding of viruses to
CC cells of a host. Oligopeptides that inhibit binding of virus to
CC receptor can be used as competitive inhibitors to release bound virus
CC in an adsorption-type assay, and if an antibody was used as an
CC immunoadsorbent, the oligopeptide could be used to elute bound virus
CC from a solid support to which virus antibody is immobilised. These
CC oligopeptides may further be used to bind to host antibody to provide a
CC transient tolerant or non-responsive state.
XX Sequence 734 AA:
Query Match 84.5%; Score 2455; DB 22; Length 734;
Best Local Similarity 83.0%; Pred. No. 1.7e-198;
Matches 443; Conservative 37; Mismatches 52; Indels 2; Gaps 2;

Db 322 EYTDNDGTTTANNLTSTVOVETDSEYQLPYVLGSANOGCLPPFPADVPMIPQYGYLTLN 381
QY 181 NSGOAVGRSSFFCYLXEPSPGMLRTGNNTFTSYEEVPPHSSYAHOSLDRLNPLIDQY 240
Db 382 NSGOAVGRSSFFCYLXEPSPGMLRTGNNTFTSYEEVPPHSSYAHOSLDRLNPLIDQY 441
QY 241 LYYINRTONQSGSAOKKDLLFSRGSFAGMSVQPKMLPGPCYRQORVSKTKTDNNNSNFT 300
Db 442 LYYLSRTNTPSGTTTQSRLOFSGAGASDLRDGRNMLPGPCYRQORVSKTSADNNSEYS 501
QY 301 WTGASKNNLNGRSTINPGTAMASHKODEKFPFPMGCVNMFEGESAGASTALDNNMTD 360
Db 502 WTGATKXHLNGRSDLVNPGPAMASHKODEKFPFPGSGVLI FGKQSEKTNVDIEKMTD 561
QY 361 EBEIKATNPVATERFGTVAVNPFSSSTD PATGTVHMGALPGWMDRDVYLGPIWAKI 420
Db 562 EBEIRTTN-VATIQYSVSINLQGRKQATADVNTQGVLPGVNMQDRDYLQPIWAKI 620
QY 421 PHTDGHFHPSPMLGFGFKHPPOILLIKTVPANPSTTFSAKFAFSTTQYSTGVSV 480
Db 621 PHTDGHFHPSPMLGFGFKHPPOILLIKTVPANPSTTFSAKFAFSTTQYSTGVSV 680
QY 481 IEMELOKENSKRNNPEVQYTSNYSKASANDFTDNNGLYTEPRIGTRYLTRL 534
Db 681 IEMELOKENSKRNNPEIQYTSNYSKASVAVDFTVDINGVSEPRIGTRYLTRL 734
RESULT 15
AAM46314
ID AAM46314 standard; Protein; 544 AA.
XX AAM46314;
XX 28-AUG-1998 (first entry)
DE AAV4 VP3 coat protein.
XX AAV; AAV-4; VP2; coat protein; vector; gene transfer;
KM gene delivery; cancer; gene therapy.
XX Adeno associated virus 4.
OS
XX Key Location/Qualifiers
FH Misc-difference 18
FT /note= "encoded by GGC"
FT /note= "encoded by CTT"
FT Misc-difference 60
XX WO9811244-A2.
XX 19-MAR-1998.
XX 11-SEP-1997; 97WO-US16266.
XX 11-SEP-1996; 96US-0025934.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Chlorini JA, Kotin RM, Safer B;
PI WPI; 1998-207403/18.
XX N-PSDB; AAV21658.
XX Adeno-associated virus 4 based vectors - used to transduce erythroid
PT progenitor cells for treatment of e.g. cancer and other genetic
PT diseases
XX Disclosure; Page 67-68; 80pp; English.
XX This polypeptide comprises the adeno associated virus 4 (AAV4)
XX VP3 coat protein that is encoded by a nucleotide sequence (see
XX AAV21658) of the AAV4 genome (see also AAV21648). The invention also
XX provides AAV4 VP1 (see AAM46308) and VP2 (see AAM46313). The invention

CC provides recombinant vectors and viral particles based on AAV4 that
 CC may be useful for transducing erythroid progenitor cells for the
 CC treatment of e.g. cancer and genetic diseases which can be corrected
 CC by bone marrow transplants using matched donors. A claimed method
 CC of delivering a nucleic acid to a subject (including a subject with
 CC antibodies to AAV2) comprises administering to a cell from the
 CC subject an AAV4 particle comprising the nucleic acid inserted
 CC between a pair of AAV inverted terminal repeats (see AAV21651 and
 CC AAV21659), and returning the cell to the subject. An antibody
 CC against AAV4 VP2 can be used to determine the suitability of an
 CC AAV4 vector for administration to a subject. AAV4 VP2 can also be
 CC used to detect the presence of AAV4-specific antibodies in a sample.
 XX

XX Sequence 544 AA;

Query Match 58.2%; Score 1690.5; DB 19; Length 544;
 Best Local Similarity 59.4%; Pred. No. 4.3e-134;
 Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;

Qy	2	ASGGGAPMADNNEGADGVGNAGNMGCDSTWLGDRVITTSRTWALPTYNHLYKQISSA	61
Db	9	AAAGGAIV-EGGQADGVGNAGNMGCDSTWSEGHVITTSRTWALPTYNHLYKRLGE-	66
Qy	62	STGASNDNHYFGYSTPWGYPDFNRFCHFSPEMDWRLINNMGFRPKRLNFKLFIQVKE	121
Db	67	---SLGNTYNGFSTPWGYPDFNRFCHFSPEMDWRLINNMGFRPKRLNFKLFIQVKE	123
Qy	122	VTTNDGVTITANNITSTVOVPSDEYQLPYVLSAHQGLPPPADVPMIPQY---LT	178
Db	124	VTTNNGETVANNITSTVOIFADSSEYELPYMDAGQESLPPFPNDVPMVQYCGIYVT	183
Qy	179	LNNQSAVGRSSFYCLEYFSPSOMLRTGNPFSTYFPEVPHSSYASOSLDRMLNPLID	238
Db	184	GNTSQOQTDNNAFYCLEYFSPSOMLRTGNPFSTYFPEVPHSSYASOSLDRMLNPLID	243
Qy	239	QYLYTLNRTON---QSGSAQNKDLFPSRGSFAGMSVQPKNWLPGCYRQQRVSKTKTDN	294
Db	244	QYLMGLQSTTGTTLNAGTATTN---FTKLKPTNFSNPKKNWLPGPSIKQGFSKTA--N	298
Qy	295	NNSNFTWTGAS--KY---NLNGRESTINDGTAMASHKDEDEKFFPMGVMIFGKESAG	347
Db	299	QNKKIPATGSDSLIKYETHSTLDRGWSALTPGPMATAGPADSK-FSNSQLIFAGPKONG	357
Qy	348	ASNTALDNVMTDEEEKATNPVATERPGTAVAVNFQSSSTDPAATGVDHAGALPGMWOD	407
Db	358	NTATVPGLTIFTSEELATNATDTDMGNDLPGDQNSNLPYVDRLTALGAVGMYWON	417
Qy	408	RDVYLQSPIMAKIPHTDGHFHPSPFLMGSGFGLKNPPQILIKNTEVPANPPAEFSATKFS	467
Db	418	RDLYQGPIMAKIPHTDGHFHPSPFLMGSGFGLKNPPQILIKNTEVPANPPAEFSATKFS	477
Qy	468	FTIOYSTGOVSVEIEMELOKENSKRANPEVOYTSNYAKSANVDFTVNNGLYTEPRPIGT	527
Db	478	FTIOYSTGOVSVOIDWEIOKERSKRANPEVOFTSNYQONSLLMAPDAGKYTEPRAIGT	537
Qy	528	RYLTRPL 534	
Db	538	RYLTRPL 544	

Search completed: July 17, 2003, 18:31:21
 Job time : 38.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 13.1429 Seconds
(without alignments)
1195.466 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGRYLTRPL 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A COMB pep:*
2: /cgn2_6/ptodata/1/1aa/5B COMB pep:*
3: /cgn2_6/ptodata/1/1aa/6A COMB pep:*
4: /cgn2_6/ptodata/1/1aa/6B COMB pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS COMB pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1690.5	58.2	544	US-09-532-594B-18	Sequence 18, Appl
2	1690.5	58.2	598	US-09-532-594B-16	Sequence 16, Appl
3	1690.5	58.2	734	US-09-532-594B-4	Sequence 4, Appl
4	479.5	16.5	543	US-08-856-841-22	Sequence 22, Appl
5	326	11.2	415	US-08-856-841-20	Sequence 20, Appl
6	318	10.9	395	US-08-856-841-13	Sequence 13, Appl
7	318	10.9	486	US-08-856-841-19	Sequence 19, Appl
8	317.5	10.9	500	US-08-856-841-16	Sequence 16, Appl
9	317.5	10.9	501	US-08-856-841-18	Sequence 18, Appl
10	304.5	10.5	264	US-08-856-841-14	Sequence 14, Appl
11	303	10.4	398	US-08-856-841-21	Sequence 21, Appl
12	209.5	7.2	579	US-09-532-594B-13	Sequence 13, Appl
13	183	6.3	584	US-09-532-594B-2	Sequence 2, Appl
14	171	5.9	387	US-08-856-841-17	Sequence 17, Appl
15	111	3.8	655	US-08-469-202-27	Sequence 27, Appl
16	111	3.8	655	US-08-484-434C-34	Sequence 34, Appl
17	109	3.8	1186	US-08-485-568A-4	Sequence 4, Appl
18	109	3.8	1186	US-08-357-658-6	Sequence 6, Appl
19	109	3.8	1186	US-08-590-554A-4	Sequence 4, Appl
20	109	3.8	1186	US-09-184-223-4	Sequence 4, Appl
21	109	3.8	1186	PCT-US93-12682-6	Sequence 6, Appl
22	108	3.7	624	US-08-947-965-78	Sequence 78, Appl
23	108	3.7	655	US-08-469-202-28	Sequence 28, Appl
24	108	3.7	655	US-08-484-434C-35	Sequence 35, Appl
25	104.5	3.6	717	US-09-626-589-1	Sequence 1, Appl
26	104	3.6	1013	US-09-415-522-8	Sequence 8, Appl
27	102	3.5	10182	US-09-134-001C-3159	Sequence 3159, Ap

28	99	3.4	824	4	US-09-626-589-3	Sequence 3, Appl
29	98.5	3.4	997	1	US-08-232-540-1	Sequence 1, Appl
30	98.5	3.4	997	1	US-08-428-949A-1	Sequence 1, Appl
31	98.5	3.4	997	1	US-08-428-949A-1	Sequence 1, Appl
32	98.5	3.4	997	2	US-08-428-946-1	Sequence 1, Appl
33	98.5	3.4	997	5	PCT-US95-04656-1	Sequence 1, Appl
34	98.5	3.4	998	1	US-08-233-008A-6	Sequence 6, Appl
35	98.5	3.4	1000	4	US-09-193-562D-30	Sequence 30, Appl
36	98.5	3.4	1021	1	US-08-233-008A-2	Sequence 2, Appl
37	98	3.4	1178	6	US254799-5	Patent No. 5254799
38	97	3.3	540	4	US-08-687-580B-7	Sequence 7, Appl
39	97	3.3	659	4	US-09-626-589-2	Sequence 2, Appl
40	96.5	3.3	1651	4	US-09-540-245A-18	Sequence 18, Appl
41	96.5	3.3	2237	1	US-08-354-973-1	Sequence 1, Appl
42	96	3.3	1177	6	US169629-2	Patent No. 5169629
43	96	3.3	1394	4	US-08-296-791-2	Sequence 2, Appl
44	96	3.3	1394	5	PCT-US95-10661A-2	Sequence 2, Appl
45	95	3.3	680	3	US-08-947-965-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-532-594B-18
Sequence 18, Application US/09532594B
Patent No. 6468524
GENERAL INFORMATION:
APPLICANT: Chortini, John A.
APPLICANT: Kotlin, Robert M.
APPLICANT: Safier, Brian
APPLICANT: Davidson, Beverly
TITLE OF INVENTION: AA4 VECTORS AND USES THEREOF
FILE REFERENCE: 14014.025202
CURRENT APPLICATION NUMBER: US/09/532,594B
CURRENT FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 544
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
NAME/KEY: misc:feature
OTHER INFORMATION: AA4 capsid protein VP3
US-09-532-594B-18

Query Match 58.2%; Score 1690.5; DB 4; Length 544;
Best Local Similarity 59.4%; Pred. No. 2.4e-146;
Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;
QY 2 ASGGAPMADNNEGADGVGASGNHCHDSITWLCDRIYTTSTRMALPTNNHLYKOISSA 61
DB 9 AAGGAAV-EGGGAGGVGASGDWHDSTWSGHTTTSTRVWLPPTNNHLYKGLGE- 66
QY 62 STASNDNHFGYSTWGFDFENRPHCHSPRDWOLINNMGPRRLNFKLPNTQVKE 121
DB 67 ---SLOSNTNGSTPWGYDFENRPHCHSPRDWOLINNMGPRRLNFKLPNTQVKE 123
QY 122 VTINDVTTANNLTSTVOFSPSEYOLPVYGSAAHOGCLPPPAVFMPOGY--LT 178
DB 124 VTISNGETVANNLTSTVOLFADSSYELFPYMDAGGSLPPFNDVFMPOGYCGSLVT 183
QY 179 LNNQSAVGRSSFYCLEPSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRIANPLID 238
DB 184 GNTSQOQTRNNAVYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSVYASOSLDRIANPLID 243
QY 239 QYIYIYANTRN---OSGSAQNDLLFSGSGPAGMVGVPNNWLPGRCYRQQRSKITDN 294
DB 244 QYIMGLOSTTGTTLNAGATATN---FTKLRLPTNPNFKNWLPGPSIKQOQSKTA--N 298

[illegible]

```

RESULT 2
US-09-532-594B-16
: Sequence 16, Application US/09532594B
: Patent No. 6468524
: GENERAL INFORMATION:
: APPLICANT: Chorini, John A.
: APPLICANT: Kotrin, Robert M.
: APPLICANT: Safer, Brian
: APPLICANT: Davidson, Beverly
: TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
: FILE REFERENCE: 14014.025202
: CURRENT APPLICATION NUMBER: US/09/532,594B
: CURRENT FILING DATE: 2000-03-22
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 16
: LENGTH: 598
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
: OTHER INFORMATION: Synthetic construct
: NAME/KEY: misc feature
: OTHER INFORMATION: AAV4 capsid protein VP2
US-09-532-594B-16

```

[illegible]

Db	353	QWKPATGSDSLIKYETHSTLDGWSALTPGPMATAGPADSK-FSNSOLIFAGPKONG	411
Qy	348	ASATLADNMIDDEBEIKATNPVARERGTAUVAFPOSSSTDPAQGDVHNAIGMWMOD	407
Db	412	NKATVPGLTFTSEEBLATNATDTDMGNI.PGGBQDSUNLPYDRLTALGAIVGMWON	471
Qy	408	RDVYLQGPIMAKIPHTDGHFHSPLMGFGKKNPPOLLKNTPEVNPAPAEFSATKFA5	467
Db	472	ROIYVGPIWAKIPHTDGHFHSPLIGFGKHPPEOLFKNTPVPANPATTFSSFPVNS	531
Qy	468	FITQYSTQGVSVIEIMELQENSKRNNEUYTSNYSKANSNVDDTYUNNGJLXTEPPRIGT	527
Db	532	FITQYSTQGVSVQIOMEIQRKSKRNNEVOFTSNYGQNSLMAAPDAAGKTEPRIGT	591
Qy	528	RYLTPRL 534	
Db	592	RYLTNHL 598	

```

RESULT 3
US-09-532-594B-4
: Sequence 4, Application US/09532594B
: Patent No. 6468524
: GENERAL INFORMATION:
: APPLICANT: Chorin, John A.
: APPLICANT: Kotin, Robert M.
: APPLICANT: Saefer, Brian
: APPLICANT: Davidson, Beverly
: TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
: FILE REFERENCE: 14014.025202
: CURRENT APPLICATION NUMBER: US/09/532,594B
: CURRENT FILING DATE: 2000-03-22
: NUMBER OF SEQ. ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 734
: TYPE: PR1
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence, No. 6468524e =
: OTHER INFORMATION: Synthetic construct
: NAME/KEY: misc feature
: OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

```

Query	Match	Similarity	59.4%;	Score 1690.5;	DB 4;	Length 724;
Best Local	Similarity	59.4%;	Pred. No. 3.8e-146;			
Matches	325;	Conservative	65;	Mismatches	133;	Indels 25; Gaps 9
QY	2	ASGGGAPADNNEGADGVGNASGNMHCDSITLADRVITTSRTRYALPTYYNHLYKQISSA	61			
DB	199	AAAGGAALV-BEGQAGADGVGNASGDHMCDSITSEGHVITTSRTRYALPTYYNHLYKRLGE-	256			
QY	62	STGASNDNHFGYSTPWPGYDPNRFPHCFHSFRDQRLINNNWGRPPKLNFKLFNIQYKE	121			
DB	257	---SLQSTNTYGFSTPWPGYDPNRFPHCFHSFRDQRLINNNWGMKPKAMRYKIFINIQYKE	313			
QY	122	VTINDGVTTIANNLSTPQVFSDEEYOLPYLIGASHQCILPPPADVPMIPQYGY---LT	178			
DB	314	VTTISNGETTYANNLITSTYQILFADSYTELPYMDAQBSLSLPPFPDIVMVMYQYGCGLVT	373			
QY	179	LNNGSOAVGRSFCYLEYFSPQMLRTGNFTFSYTFEEVPFHSYAHSQSLDLRLNNPLID	238			
DB	374	GNTSQQOQDRNAFYCLEYFSPQMLRTGNFETIYSFEKVPFHSYAHSQSLDLRLNNPLID	433			
QY	239	QYLYLAFRTQN---QSSGAQNKLLFSRSGSPAGMSVQPKWMLGQCCTCRQORVSTTKTDN	294			
DB	434	QYLMGLQSTTYGTLINAGTATN---PTFKARPTFNSPKKMWLQGPSLKQOGFSKTA--N	488			
QY	295	NNNSFWTWTGAS---KY-----NLNRESITINEGTAMASHKDDQEDKFPMSGVIPEKESAG	347			
DB	489	QNYKIPATGSSLIKETHTSTLDRKWSMLTPGPPMATAIGPADSK-FNSQILIFACPKONG	547			


```

QY 348 ASNTALDNNWITDEEERIKATNPVATERFTVAIVNROSSSTDPATCDVAMGALPGMTWOD 407
Db 548 NTATVETGTLFTSEELAAATNATDIDMGANLPCGQOSNINPYDRLTRALAAVGMWON 607
QY 408 RDVYLGGPIWAKIIPHTDGHFHPSPLMGFGGLKNPPQILIKNTPEVNPANPAEFSATKFA5 467
Db 608 RDIIYGGPIWAKIIPHTDGHFHPSPILGGFGGLKHPPIQIFIKNTPEVNPANPATFTSTVNS 667
QY 468 FTIYVSTGVSVIEIHELQENSKRNPEVQTSYNAKSAVDFVNNGLYTEPRIGT 527
Db 668 FTIYVSTGVSVQIDWEIQERSKRNPEVQTSYNGQNSILWAPDAAGKYTEPRAIGT 727
QY 528 RYLTRPL 534
Db 728 RYLTRHL 734

RESULT 4
US-08-856-841-22
Sequence 22, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED NOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038562
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 543
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE

```

```

/ POSITION IN GENOME: N/A
/ FEATURE:
/ NAME/KEY: N/A
/ LOCATION: N/A
/ IDENTIFICATION METHOD: amino acid analysis and
/ IDENTIFICATION METHOD: mass spectrometry
/ OTHER INFORMATION:
/ PUBLICATION INFORMATION:
/ AUTHORS: COSSART, Y.E.
/ AUTHORS: FIELD, A.M.
/ AUTHORS: CANT, B.
/ AUTHORS: WIDDOWS, D.
/ TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
/ JOURNAL: LANCET
/ VOLUME: 1
/ ISSUE:
/ PAGES: 72 - 73
/ DATE: 1975
/ DOCUMENT NUMBER:
/ FILING DATE:
/ PUBLICATION DATE:
/ RELEVANT RESIDUES IN SEQ ID NO: 22:
US-08-856-841-22

Query Match 16.5% Score 479.5; DB 4; Length 543;
Best Local Similarity 27.1%, Fred. No. 1.9e-35;
Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16

QY 10 ADNNEGADGVGNAS--GNMHCDSTWLGDRVITTSRTTALPTYNHLYKQISSASTGASN 67
DB 16 AEASITGAGGGGSSNYSKMSSEGAITFSANSVCTPESROGLIYPDEHHIKVYSSPAASSCHN 75
QY 68 D-----NHFGYSTPWGYDFDNFCHGFSPEDMQRLLINNNGRRPKRLNKLFINI 117
DB 76 ASGREAKVCITSPIMGYSTPWRYLDFNALNLFSPLEFQHLIENGVSIAPDALVTITSEI 135
QY 118 QYKEVT--TNDGYTTIANNLITQYVESDSYOLPYVYLSAAGCLPFPADVFMIPOYG 175
DB 136 AVKQVTDKRTGGV-QVTDSTTGRLCMLVDHEYKPYVYGGOGDTLAPLPIWVYPPQYA 194
QY 176 YLTINN-GSAVAG-----RSFSYCLEYFPSSOMLTGANNFPFSYTFEEVPHFSYAH 225
DB 195 YLTGVDVNTQGISGDSKCLASEESAFYVLEHSSFQLTGTGTAWSYKFPVPVPPENLEG 254
QY 226 SOSIDRLNPLIDQYLYIANRTONOSGSAQMKDLLFSRGSFAGMSVQPKNLLPGPCYRQ 285
DB 255 SQHFYEMYNPL---YGSRLGVPDILGDPKFRSL-----THEBIAIQQNTMGPGLVNSV 306
QY 286 RVSKTKTDNNNSNFTWGASKYNNLNGRESINPG-TAMASHKDDKCFEPMGVMNIPEKE 344
DB 307 STKSGDSSNTGAKALTELSTGTSQNTREISLRPGVSPQRYHHMDPDKVYTGAINAISHQ 366
QY 345 SAGASNTALDNV-----MTDDEELIATNPVATERGCIYAVNPQSSSTDPATCDVAM 397
DB 367 TYG--NAEDREYQGVGRFPEKEKQLQOLGLNNHTY-----FPNGTQYTDQIR-R 416
QY 398 GALPGMWQODRDVYLOGEITAKIPIHTDGHFAPS-PLMGFGILKPNPOLLIKNTVPANP 456
DB 417 PLMGVSNVRRLALHYESQLSKINLSDSFTQCPAALGCKWGLHQDPPI-----465
QY 457 PAEFSATKFAFTIYQSTGVSVLEIWEEL-QKENSKRANPE 496
DB 466 -----FLKQYAVGIMVTMTFELGPRKATGRANPQ 495

RESULT 5
US-08-856-841-20
; Sequence 20, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: ERWIN SOUTSCHEK
; APPLICANT: MANFRED MOTZ
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES

```

TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 567-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 415
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.B.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE: 72 - 73
PAGES: 1975
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 20:
US-08-856-841-20

Query Match 11.2%; Score 326; DB 4; Length 415;

Best Local Similarity 26.0%; Pred. No. 1,4e-21;
Matches 101; Conservative 66; Mismatches 180; Indels 42; Gaps 13;
QY 128 VTTIANNLSTVQFSDSEYQLPVIGSAHQGLPPADVEMIPQYGYITLNN-GSQAV 186
Db 1 MTWITSLHACMLV--DHEKYKPYVIGQCGDPLAPELPIWVYPPQYAYLTGADVNTQGI 58
QY 187 G-----RSSFCLEYRPPSQMLRTQNNFTFSYEEBEPFSSVAHSQSLRLNPLI 237
Db 59 SGDSKLIASESAFYVLEHSSFQGLTGTAISMSTYFPVPENLEGCSSHFTYEMNPL- 117
QY 238 DQVLYLNRTONQSGANQDLFSRSGSPAGMSVOPKMLPGCYRQORVSKTKTNNNS 297
Db 118 --YGRAGVPDITGQDPKPSL-----THEDHAIQPNFMRGPLVNSVSTKEGDSNTGA 170
QY 298 NFWTGAASKYNLNGRESITPG-TANASHKDDKDFPMSGVMIFGKESAGANTALDNV 356
Db 171 GKALTGISTGTSQNTSRISLPGFVSPYHHMDTKVTGAINASHQYTYG--NADKE 227
QY 357 -----MTDEEIKATNVATERFGTVAVNFQSSSTDPATGDAVMGALPGMVMQDRD 409
Db 228 YQGVGRFPEKQQLQLOGLNHTY-----FNNKGTQYTDQIF-RPLMVSVMNRRA 280
QY 410 VYLQGPIMAKIPHTDGHFHS-PLMGSGFLKNPPOILIKNTFVPANPAPFSGATKFSF 468
Db 281 LHYESQUMSKIPNLDSFKQFALGSGWGLHQPPQIFLK--ILPSGPIGIGKSMGIT 358
QY 469 ITQYSTGQVSEIEMEL-QKNSKRNPPE 496
Db 339 LVQYAVGIMVTMTFGLGPRKATGRWNPQ 367
RESULT 6
US-08-856-841-13
Sequence 13, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUNSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS: 28
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355

TELEFAX: (212) 557-5635
 TELEX: NONE
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 395
 TYPE: AMINO ACID
 TOPOLOGY: LINEAR
 MOLECULE TYPE:
 DESCRIPTION: PEPTIDE
 HYPOTHEICAL: N/A
 ANTI-SENSE: N/A
 FRAGMENT TYPE: INTERNAL
 ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
 ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)
 IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
 POSITION IN GENOME: N/A
 FEATURE:
 NAME/KEY: N/A
 LOCATION: N/A
 IDENTIFICATION METHOD: amino acid analysis and
 IDENTIFICATION METHOD: mass spectrometry
 OTHER INFORMATION:
 PUBLICATION INFORMATION:
 AUTHORS: COSSART, Y. E.
 AUTHORS: FIELD, A. M.
 AUTHORS: CANT, B.
 AUTHORS: WIDDOWS, D.
 TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
 JOURNAL: LANCET
 VOLUME: 1
 ISSUE:
 PAGES: 72 - 73
 DATE: 1975
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO: 13:
 PUBLICATION INFORMATION:
 AUTHORS: MANIATIS, T.
 AUTHORS: FRITSCH, E. F.
 AUTHORS: SAMEROOK, J.
 TITLE: MOLECULAR CLONING
 JOURNAL: COLD SPRING HARBOR, NY
 VOLUME:
 ISSUE:
 PAGES:
 DATE: 1982
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO: 13:
 PUBLICATION INFORMATION:
 AUTHORS: SMITH, D. B.
 AUTHORS: JOHNSON, K. S.
 TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
 TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
 TITLE: GUTATHIONE S. TRANSFERASE
 JOURNAL: GENE
 VOLUME:
 ISSUE: 67
 PAGES: 31 - 40
 DATE: 1988
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO: 13:
 US-08-856-841-13

Query Match 10.9%; Score 318; DB 4; Length 395;
 Best Local Similarity 25.4%; Pred. No. 7e-21;
 Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;
 128 VTIANNLTGTVGVSDSEYQLPYVLSAHQGLPPPADVFMIQYGYLTINN-GSQAV 186

Db 1 MMTITPSLHACMLV--DHEKYPYVLGGQDDTLAPELPIVWYPPQYAYLTVDVNTQGI 58
 Qy 187 G-----RSSFYCLEYFPSSQMLTGNFTSYFEEYFPFSSVAHSQSLRLNPLI 237
 Db 59 SGDSKSLASESAFYVLEHSSFQLLTGTASMSYKFPVPFNNLBCSOFYEMYNPL- 117
 Qy 238 DQYLYLNTQNGSGAQNNDLFSRSGPAGMSVOPKNNLPGCYQQRASKTKTNNNS 297
 Db 118 --YGRLVGVDTLGGPKFRSL-----THEDHAIQPNFPGPLVNSVSTKEDSSNTGA 170
 Qy 298 NPTWTGASKYNLNGRSIINPG-TAMASHKDEDEKFPFMSGVMIFGESAGANTLDNV 356
 Db 171 GKALTLSTSGNTSLRPGVSPYHHMDTKVVTGINALSHGQTYG---NAEDKE 227
 Qy 357 -----MTDEEITATNPVATERGTVAANFQSSSTBDPATGVAMGALPGMWNQDRD 409
 Db 228 YQGVGRFPNEKEQLKQLQGLNHTY-----FPNKGTOQYTDQIE-RPLMVGSVNNRRA 280
 Qy 410 VYLGSPIMAKIPHTDGHFHS-PLMGFGIKNPPQILIKNTVPANPPAFSATKPSF 468
 Db 281 LHTESQLSKIPVLDSEFKTQFALGSGWGLHOPPQI-----F 318
 Qy 469 ITQYSTGVSVETEMEL-OKENSKRNPNP 496
 Db 319 LKQYAVGIMVTMTFTLGPRAKATGRNPPQ 347

RESULT 7
 US-08-856-841-19
 Sequence 19, Application US/08856841
 Patent No. 6274307
 GENERAL INFORMATION:
 APPLICANT: ERWIN SOUTSCHEK
 APPLICANT: MANFRED MOTZ
 TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
 TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
 NUMBER OF SEQUENCES: 28
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
 STREET: 99 PARK AVENUE
 CITY: NEW YORK
 STATE: NY
 COUNTRY: USA
 ZIP: 10016
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" FLOPPY DISC
 COMPUTER: AT&T - IBM COMPATIBLE
 OPERATING SYSTEM: MS-DOS Version 6.2
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/856,841
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/214,658
 FILING DATE: 16-MARCH-1994
 APPLICATION NUMBER: US 07/917,096
 FILING DATE: 4-AUGUST-1992
 APPLICATION NUMBER: PCT/DE91/00106
 FILING DATE: 8-FEBRUARY-1991
 APPLICATION NUMBER: DE40038262
 FILING DATE: 8-FEBRUARY-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: ROBINSON, WILLIAM R.
 REGISTRATION NUMBER: 27,224
 REFERENCE/DOCKET NUMBER: LKR-9222-A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-3355
 TELEFAX: (212) 557-5635
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:

LENGTH: 486
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 19:
US-08-856-841-19

Query Match 10.9%; Score 318; DB 4; Length 486;
Best Local Similarity 32.8%; Pred. No. 9.6e-21;
Matches 82; Conservative 31; Mismatches 109; Indels 28; Gaps 6;

QY 14 EGAGGVGASGN-----WHCSTMLGDRVITTSRTMALPTVNNHLYQJISASTGASND 68
DB 235 EASTGAGGGGSPVSKMSWSEATPSANVCTFSQFLIPDPPEHYVVFSPASSCHNA 294
QY 69 -----NHFGYSTPGVGFDFNRFCHFSPPDMQRLINNNGFPRKLNFKLFNIQ 118
DB 295 SGKEAKVCTTIPINGYSTPMRYLDPMNALNLFSPLEPHLENGSTPBDLTYTISIA 354
QY 119 VKEYT--TNDGVTTIANLSTVOVSDSEYQLPYVLSAHQGLPPPADVFMIPQY 176
DB 355 VKDVTDKTGGGV-QVTDSATGRLCLVDHEVKYPVVLGQGDITLAPBLPIWVYPPQYAY 413
QY 177 LTLNN-GSOAVG-----RSSPYCLEYPPSQMLRTGNNTFSYTEEVPFHSVAHS 226
DB 414 LTVGVNTQGISGDSKSLASBSAPYVLEHSSFOLLGTGTMTSYKPPVPPENLBGS 473
QY 227 QSLRLMNP 236
DB 474 QHFYEMNPL 483

RESULT 8
US-08-856-841-16
Sequence 16, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 90 PARK AVENUE

CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 500
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: I
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:

PAGES: 1982
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME: 67
ISSUE: 31 - 40
PAGES: 1988
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 16:
US-08-856-841-16

Query Match 10.9%; Score 317.5; DB 4; Length 500;
Best Local Similarity 33.1%; Pred. No. 1.1e-20;
Matches 83; Conservative 34; Mismatches 109; Indels 25; Gaps 6;

QY 10 ADNNEGADGVGNAS--GNHCDSTWLGDRVITTSRTWMLPTNNHLYKOISSASTGASN 67
DB 242 AEASTGAGGGGSSVSKMSWSEGATFSANSVCTFSRQFLIPYDEHHYKVFSPAASSCHN 301
QY 68 D-----NHFGYSTPMGYDFNRFCHPSPDMORLINNNWGFRPKRLNFKLFNI 117
DB 302 ASGEKAVCTISPMGISTPWRYLDPNALNLFSPLEFOLHLENYGSIAPDALTVITISEI 361
QY 118 QVKEVT--TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHOGCLPPFPADVFMIPQY 175
DB 362 AVKDVTDKTGGGV-QVTDSTTGRCLMLVDHEYKYPVVLGGQODTLABELPIWVFPPOYA 420
QY 176 YLTIAN-GSQAVG-----RSSFYCLEYFSPQMLRTGNFTFSYTEEVPFHSYAH 225
DB 421 YLTVDVNTGIGSDSKKASBSAFYVLBRHSSPOLLGTGTASMSYKFPVPPENLEGC 480
QY 226 SOSLDRLMNP 236
DB 481 SOHRYEMNPL 491

RESULT 9
US-08-856-841-16
Sequence 16, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYPOTHEICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 18:
US-08-856-841-16

Query Match 10.9%; Score 317.5; DB 4; Length 501;
Best Local Similarity 33.1%; Pred. No. 1.1e-20;
Matches 83; Conservative 34; Mismatches 109; Indels 25; Gaps 6;
QY 10 ADNNEGADGVGNAS--GNHCDSTWLGDRVITTSRTWMLPTNNHLYKOISSASTGASN 67
DB 243 AEASTGAGGGGSSVSKMSWSEGATFSANSVCTFSRQFLIPYDEHHYKVFSPAASSCHN 302
QY 68 D-----NHFGYSTPMGYDFNRFCHPSPDMORLINNNWGFRPKRLNFKLFNI 117
DB 302 ASGEKAVCTISPMGISTPWRYLDPNALNLFSPLEFOLHLENYGSIAPDALTVITISEI 362
QY 118 QVKEVT--TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHOGCLPPFPADVFMIPQY 175
DB 363 AVKDVTDKTGGGV-QVTDSTTGRCLMLVDHEYKYPVVLGGQODTLABELPIWVFPPOYA 421
QY 176 YLTIAN-GSQAVG-----RSSFYCLEYFSPQMLRTGNFTFSYTEEVPFHSYAH 225
DB 421 YLTVDVNTGIGSDSKKASBSAFYVLBRHSSPOLLGTGTASMSYKFPVPPENLEGC 480

Db 422 YLTVGDNVNTQGISGDSKSLASESAFYVLEHSSFOLLGTGTASMSYKFPVPENLEG 481
QY 226 SGLSDRLMNP 236
Db 482 SQHFYEMYNPL 492

RESULT 10

US-08-856-841-14
Sequence 14, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856, 841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214, 658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917, 096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 264
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: N/A
HYDROTHERICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.

AUTHORS: CANT, B.
AUTHORS: WIDOMAS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMBROOK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH
TITLE: GLUTATHIONE S. TRANSFERASE
JOURNAL: GENE
VOLUME:
ISSUE: 67
PAGES: 31 - 40
DATE: 1988
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 14:

US-08-856-841-14
Query Match 10.5%; Score 304.5; DB 4; Length 264;
Best Local Similarity 32.5%; Pred. No. 6.5e-20;
Matches 76; Conservative 30; Mismatches 105; Indels 23; Gaps 5;

QY 7 APNADNNEGADGVGNASGNHCDSTWLGDRVITTTSTRTMALPTYNHLYQISSASTGAS 66
Db 9 AAEASTGAGGGGSGNSVKSMSBEGATFSANSVCTFSQFLLPYDPEHHYKVFSPASSCH 68
QY 67 ND-----NHFGYSTPMGYDPENRFRHCHSPRDMQRLINNNMGFRPKLNFLEN 116
Db 69 NASGKRAKCTTISPINGYSTPMRYLDPNALNLFSPLEFQHLLENYSINPDALITYTISE 128
QY 117 IOVEKVT--TNDGVTTIANLSTVOYFSDSEYQLPYVLSAHOGLPPPADVFMIPQY 174
Db 129 IAVKDVTDKTKGGV-QVTDSTGRCLMLVDHEKYPYVLGGQDTLAPSLPIWVYFPQY 187
QY 175 GYLTLNN-GSQAVG-----RSSFYCLEYFSPQMLRTGNNFTSYTEEVP 218
Db 188 AYLTVGDNVNTQGISGDSKSLASESAFYVLEHSSFOLLGTGTASMSYKFPVP 241

RESULT 11

US-08-856-841-21
Sequence 21, Application US/08856841
Patent No. 6274307
GENERAL INFORMATION:
APPLICANT: ERWIN SOUTSCHEK
APPLICANT: MANFRED MOTZ
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
8

```

Oy      14  USEQLPVLGSAHQGLPPRADVFMIPQYGYLTIAN-GSOAVG-----RSSFYC 193
Db      4  DHEKYPVLVGGODDTLAPELPIWVPPQYAVLTVGVNTQIGSGSKLASESAFYV 63
Oy      194 LEVPSQMLRTGNNTFSTYEEVPEFFHSVAHSOSLDLNPILDOVLVYINRTQOSGS 253
Db      64 LEHSFQLGTGGRATMSYKRPVPEPELBECOSHFYEMNPL--YGSRLGVPTLGGD 120
Oy      254 AQNKDLFSRGSPLMGSVQPKMLPGPCYRQQRVSKTTDNNNSNFTWTGASKTNLNGE 313
Db      121 PKFRSL-----THRDHAIQPNFMPGPVLVNSVSTEGDSSNTGAKALTLSTGTSQNT 175
Oy      314 SIIRPG-TAMASHODEDKFPEPMGSVMIFGESAGASTALDNV-----MITPEEIK 365
Db      176 ISLRPPSPQYHHMDTDKYTGAINAISHGQTTG--NAEDKEYQGVGFPEKEQVK 232
Oy      233 QLGQILNHTY-----PKNKGTQGYTDQIE-RPLVSGSVMRRLAHYESQLSKIPNLD 285
Db      426 HFHS-PLMGFGGLKNPPQLIKTPVNPANPAPFSATKFSFTTQYSTGVSVSEIWE 484
Oy      286 SFKTQFALGSGWGGLHQPPI-----FCYKHVQQLFVLNOME 323
Db      485 LQKNSK-----RNP 495
Oy      324 LLPNMPRELQHLNMGD 341

RESULT 12
5223424-13
Patent No. 5223424
APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
RICHARD D.
TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
ACID SEQUENCE
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/225,032
FILING DATE: 27-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 78,519
FILING DATE: 27-JUL-1987
APPLICATION NUMBER: 933,107
FILING DATE: 20-NOV-1986
APPLICATION NUMBER: 902,887
FILING DATE: 02-SEP-1986
APPLICATION NUMBER: 887,140
FILING DATE: 17-JUL-1986
APPLICATION NUMBER: 823,102
FILING DATE: 27-JAN-1986
APPLICATION NUMBER: 773,430
FILING DATE: 06-SEP-1985
SEQ ID NO:13
LENGTH: 579
5223424-13

Query Match 7.2%, Score 209.5, DB 6, Length 579,
Best Local Similarity 22.7%, Pred. No. 1,le-10;
Matches 138; Conservative 80; Mismatches 237; Indels 153; Gaps 32;

Oy      3  SGGGAPMADNNNEGAGVGNASGNMH--CDSTWIGD---RVITTSRTWALPTYNHLYKO 57
Db      24 SGGGGG-G-GGGGRGAGGVSTGTFNNTQETFOYLGBGLVRIIAHHSRLIHLNMPHEHYTKR 82
Oy      58 IS--SASTGAS---NNNHVFGYSTPGVYDPFNRFHGFSPRDWQRLINNNGWGFPRKELN 111
Db      83 IHLVNSSGVAWGAVQDDAHTQWTPSLIDRNAMGWFPNADWQLSNNTTEILVSEFE 142
Oy      112 FKLFNIOVKEVT---TNDGVTTIANLTSTVQVSDSEYQLPYVLGSAHQGLPPRADV 168

```

```

Db      143 QEIFNVVLTITTESATSPSKIYNNDLTASLWALDNTNLTPTPAARSETLGFYPMLP 202
Qy      169 FMIPQGYL-----TLNNGSQAVERG-----SFYCLE-YFPGSMLRTGNNF 208
Db      203 TKPTQYRYVYLSCLNNLPPTTGTGSOQOITDSIQGLSHDIMEFYIENAVPHILRTGDEF 262
Qy      209 -TFEYTFEEVPHSSVYASQSGLRLMNPILIDQYLYLINRTQNGSQAQNKDLFESRSPA 267
Db      263 STGIYHDTKPL-KLTHSMQTNR-----SLGLPP 290
Qy      268 GMSVQP-----KNMLPGPCYRQ--QVSKTYTD-----NNSNFTWTGAS 305
Db      291 KVLTEPTTEGDPHGHTLPGANTRKGYHOTINNSYTEATALPQAQVYTPYMNDEYNGG 350
Qy      306 KYNLNGESTINP--GTAMSHKDEDEKFPMSGVMTFGKESAGASTALDNTWITDEEE 363
Db      351 PF-----LTPVFPADQYQYDE-----PNGAIRFTMGYQHHLTTS-----SOLE 392
Qy      364 IKATNP-----VATERFGTVA-VNFOSS-----TDPATG--DVHAMGALP-----GM 403
Db      393 RYTNPQSKCGRAKQCPNQAPLNLENTNNGTLLPSPIGKSNKHPMNTLNTYGLTA 452
Qy      404 WMQDRDYVLOGPIWAKIPHTD--GHFHPSPIMGFGFLK-NPPQILIKNTFVPANPAEF 460
Db      453 LNTTAPVFPNGQIMDKELDTLKPRLH---VTAFVFCXKNPQQLFVYIAP---NLTDDE 506
Qy      461 SA--TKRASFTQSTGVSEIEMELQENSKRMNPEVOYTSYVAKSANDFTVNDGL 518
Db      507 NADSPQOPRIITD-SNFWMKGTLLFTAMRSSNMNPIQOHTT-----TAENIRK 555
Qy      519 YTEBRPIG 526
Db      556 YI-PTNIG 562

RESULT 13
US-09-022-949-2
; Sequence 2, Application US/09022949
; Patent No. 6187759
; GENERAL INFORMATION:
; APPLICANT: Tarpey, Ian
; APPLICANT: Greenwood, Neil
; TITLE OF INVENTION: Canine parvovirus DNA vaccination
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Akzo No. 6187759el Patent Dept.
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (HPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/022,949
; FILING DATE: 11-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gornley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-948-7400
; TELEFAX: 301-948-9751
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

US-09-022-949-2
Query Match      6.3%; Score 183; DB 4; Length 584;
Best Local Similarity 20.4%; Pred. No. 2.9e-08;
Matches 130; Conservative 91; Mismatches 223; Indels 192; Gaps 33;

Qy      5 GCAPMADNNEGADGVNAGSNMHCSTWLDRIYTTST-----RTWALPTYNN- 52
Db      10 GGCP-AVRNERATGSGNGSGGGGGS--GGVGISGTGFNNQTEFLENGWEITANSS 66
Qy      53 ---HL-----YKQI-----SSASTGASDNHIFGISTPMGYDFFRFGHPSRDP 95
Db      67 RLVLHNPSESENRRVYVNNLDITAVNGNALDDTAQIYTPMSLVDAVAMGVWENFGDW 126
Qy      96 QRLNNMGRPRRLNFKNIQVKEVT---TNDGVTIANLSTQVPSDSEYOLPYV 152
Db      127 QLVNTMSELHLVFEQELNVVLKTVSESAIOPPKYVYNNDLTASLWALDNNMFTT 186
Qy      153 LGSAGQCL-----PPFPA-----DVFMIPQGYLTLNNGSQAVERG----- 189
Db      187 PAMRSETLGFYPMKPTIPFWRVYFQWDRLLPSH-----TGTSGTPTNIYHGTDPPD 240
Qy      190 -SFYCLE-YFPGSMLRTGNNF-TFSTFEEVPHSSVYASQSGLRLMNPILIDQYLYINR 246
Db      241 VQFYTIENSVPVHLRTGDEGATGFEFFDCKP--CRLLTHWQTNRALG--LPFLNSLPQ 256
Qy      247 TQNGS-----GSAQNK-----DLFSRGSFAGMSVOPKNMLPGCYRQQRV 287
Db      297 SEGGINFGYIGVQDDRRGVTQMGNTNYITETALIMPAEYGS-----AYISFEAS 348
Qy      288 SK-----TKTDNNNSNFTWTGASKY--NLNGRESIINPGT---AMASHKD 327
Db      349 TQGPFTPIAARGAQIDENQAA---DGDPRYAFGRHQKQTTTGETPERFTYIAHOD 405
Qy      328 DEDKFPMSGVMTFGKESAGASTALD-NMITDEBEIKATNVVATERFTVAVNPGSS 386
Db      406 T-----GRPEEDMTQINIFNLPTNDVLLPTDPIG---GKTGINY--TN 446
Qy      387 TDPATGVDHAMGALPGWMDRDVYLOGPIWAKIPHTDGHFHPSPIMGFGFLK----- 439
Db      447 IFNTYGLTLNINVP-----PYRNGQIMDEFTD-----LKPRLHVA 486
Qy      440 -----NPPQILIKNTFVPAN--PPAEFATKFAFTQYSTGVSEIEMELQENK 490
Db      487 PFVQNNCPGQLFVKAVALNTNEYDPDASAMGR-----IYTSDFPMWKGLVFKAKLRAS 542
Qy      491 KRNNPVOYTSNYSKASANDFTVNDNGLYTERPIG 526
Db      543 HTWNPQQMSIN-----IDNQPNTV-PSNIG 567

RESULT 14
US-08-856-841-17
; Sequence 17, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: MANFRED NOTZ
; APPLICANT: ERWIN SOUTSCHEK
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
; STREET: 99 PARK AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISC
; OPERATING SYSTEM: AT&T - IBM COMPATIBLE
; SOFTWARE: ASCII

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,658
FILING DATE: 16-MARCH-1994
APPLICATION NUMBER: US 07/917,096
FILING DATE: 4-AUGUST-1992
APPLICATION NUMBER: PCT/DE91/00106
FILING DATE: 8-FEBRUARY-1991
APPLICATION NUMBER: DE40038262
FILING DATE: 8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-5635
TELEX: NONE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 387
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: INTERNAL
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: N/A
LOCATION: N/A
IDENTIFICATION METHOD: amino acid analysis and
IDENTIFICATION METHOD: mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: COSSART, Y.E.
AUTHORS: FIELD, A.M.
AUTHORS: CANT, B.
AUTHORS: WIDDOWS, D.
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL: LANCET
VOLUME: 1
ISSUE:
PAGES: 72 - 73
DATE: 1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 17:
PUBLICATION INFORMATION:
AUTHORS: MANIATIS, T.
AUTHORS: FRITSCH, E.F.
AUTHORS: SAMEROK, J.
TITLE: MOLECULAR CLONING
JOURNAL: COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE: 1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 15:
PUBLICATION INFORMATION:
AUTHORS: SMITH, D.B.
AUTHORS: JOHNSON, K.S.

```

[illegible]

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-469-202-27

Query Match 3.8%; Score 111; DB 1; Length 655;
Best Local Similarity 18.6%; Pred. No. 0.14; Mismatches 229; Indels 214; Gaps 32;
Matches 119; Conservative 77;

QY 8 PMADNNEGADGVGNAS--GNMHCSTWLGDRV-----ITSTRYVALPTYNHLYKQISS 60
DB 104 PPIDNVNTTDAAGTGYGWDYPRIDBFGNLDPKELTSLMHSPPDYMKLVLDYAP 163
QY 61 ASTGASDNRY-----FGYSTPWGYPDFNRFCHPSRDMQRLINNWG 104
DB 164 NNSNANDENEFGALYRDGVFTIDYPTDVAANTGW-----YHNGGVTNW-----NDF 210
QY 105 FRPRRLNPKLFNIQVKEVTNDGVTTANNLTSTVQVPSDEYQLPYVLGSAH--QGCLP 162
DB 211 FQVK--NHNLPNLS---DLNOSTDYQYLLDSKFWIDAGVDAIRIDAIKMDKSFQ 264
QY 163 PFPADVPMIPQYGYLTINNGSQAVGRSSFYCL-EYPSQMLRT---GN----- 206
DB 265 KWTSDI-----YDY-----SKSIGREGFPFGEMFGASANTTTGVDGNADIVANTSGSA 313
QY 207 --NFTPSYTPREVEFHSSYHNSQSLRLMPLIDQYLYL----- 244
DB 314 LLDGFPRDTLERYVGRS--GNTMKTLSYLIKQTVFTSDMDQVVFMDNMDMARIGTA 370
QY 245 ---NRT-----QNGSGAKNOLLFSR---GSPAGMSYQPKNMLPGPCYRQQRVSKTKTD 293
DB 371 LRSNATTFPGANNETGSGSQAFAQKRIDGLVATMYR---GIPATYGTGHEYAANPTS 427
QY 294 NNNNSFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPM----- 335
DB 428 N--SFGQVGSDDPYN---REKM--PGF-----DTESEAFSIIKTLGLRKSPAIQNGT 473
QY 336 -----SGVMIPGKESA-----GASNTA-LDNVMTD-----EEIKATNP 369
DB 474 YTELWVNDLILVFERRSGNDIIVIALNKGAEANTINVKNIAYENGVPBLLGNNSVSANK 533
QY 370 VAT-ERFGTVAVNFQSSSTDPAQDVHAM--GALPGWVQDRDYLOGPIWAKIPHTDGH 426
DB 534 QATITLMQNEAVVIRSQSDDAENPTVQSINPACNNGTISGQSVYIIGNI----- 583
QY 427 FHPSPLMGGFGLKNPPQILIKNTPPVPANPPAEPFSAITKFASTIYSTGVSVR----- 480
DB 584 ---POLGMDLTK-----AVKISP-----TQYPQWSASLELPSDLN 616
QY 481 IEMELQKENSGRKNPEYQYTSNYSKASANVDFTVNDNGLY 519
DB 617 VEMKCVKRNETNPTANVEMQSGANNQFNSNDTQTNSGF 655

Search completed: July 17, 2003, 18:36:39
Job time: 14.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 25.7143 Seconds
(without alignments)
2466.245 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906
Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGRYLTRPL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep1:*
11: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep2:*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep3:*
13: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2481.5	85.4	533	US-10-038-972A-15	Sequence 15, Appl
2	2481.5	85.4	533	US-10-038-972A-14	Sequence 14, Appl
3	2481.5	85.4	735	US-10-038-972A-13	Sequence 13, Appl
4	2481.5	85.4	735	US-10-293-478-1	Sequence 1, Appl
5	491.5	16.9	756	US-10-205-942-4	Sequence 4, Appl
6	108	3.7	5795	US-09-815-242-12610	Sequence 12610, A
7	104.5	3.6	717	US-10-046-583A-1	Sequence 1, Appl
8	104	3.6	2344	US-09-815-242-12713	Sequence 12713, A
9	103.5	3.6	1016	US-09-815-242-5845	Sequence 5845, Ap
10	103.5	3.6	1265	US-10-198-070-69	Sequence 69, Appl
11	102	3.5	3712	US-10-108-605-103	Sequence 103, App
12	101.5	3.5	439	US-09-836-353A-145	Sequence 145, App
13	101.5	3.5	439	US-09-984-130-145	Sequence 145, App
14	101	3.5	1411	US-10-080-505-17	Sequence 17, Appl
15	100.5	3.5	1714	US-10-098-916A-11	Sequence 11, Appl
16	99	3.4	824	US-10-046-583A-3	Sequence 3, Appl

17	98.5	3.4	1000	US-10-055-412B-30	Sequence 30, Appl
18	97.5	3.4	685	US-09-745-763-11	Sequence 11, Appl
19	97	3.3	659	US-10-046-583A-2	Sequence 2, Appl
20	97	3.3	1844	US-10-242-056-53	Sequence 53, Appl
21	97	3.3	2504	US-09-817-514A-8	Sequence 8, Appl
22	97	3.3	2504	US-10-242-056-12	Sequence 12, Appl
23	96	3.3	1394	US-09-839-996-2	Sequence 2, Appl
24	96	3.3	1394	US-10-080-505-2	Sequence 2, Appl
25	95.5	3.3	1286	US-10-153-668-212	Sequence 212, App
26	95.5	3.3	1354	US-10-153-668-470	Sequence 470, App
27	95	3.3	433	US-09-859-888-4	Sequence 4, Appl
28	95	3.3	478	US-09-859-888-2	Sequence 2, Appl
29	95	3.3	1545	US-09-839-996-4	Sequence 4, Appl
30	95	3.3	1545	US-10-080-505-4	Sequence 4, Appl
31	95	3.3	2150	US-10-135-322-17	Sequence 17, Appl
32	94.5	3.3	462	US-10-127-032-101	Sequence 101, App
33	94.5	3.3	831	US-09-732-665-7	Sequence 7, Appl
34	94	3.2	1395	US-10-080-505-7	Sequence 7, Appl
35	94	3.2	4590	US-10-160-758-13	Sequence 13, Appl
36	94	3.2	4590	US-10-160-758-14	Sequence 14, Appl
37	94	3.2	4590	US-10-060-036-157	Sequence 157, App
38	93.5	3.2	1156	US-09-826-660-15	Sequence 15, Appl
39	93.5	3.2	1178	US-09-851-194-2	Sequence 2, Appl
40	93.5	3.2	1770	US-09-841-132-444	Sequence 444, App
41	93.5	3.2	1848	US-09-839-996-6	Sequence 6, Appl
42	93.5	3.2	1848	US-10-080-505-6	Sequence 6, Appl
43	93	3.2	1404	US-09-811-045A-1	Sequence 1, Appl
44	92.5	3.2	716	US-09-349-058-38	Sequence 38, Appl
45	92.5	3.2	716	US-10-040-430-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-10-038-972A-15
; Sequence 15, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/3696US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260, 124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 533
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP3 capsid protein
; US-10-038-972A-15

Query Match	85.4%; Score 2481.5; DB 15; Length 533;
Best Local Similarity	83.3%; Pred. No. 7.5e-225;
Matches	445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;
QY	1 MASGGAPMADNNEGADGVGASGNHCDSTWIGDRVITSTRTWALPTYNNHLYKQISS 60
DB	1 MATSGAPMADNNEGADGVGSSGNHCDSTWIGDRVITSTRTWALPTYNNHLYKQISS 60
QY	61 ASGASNDNHVPCYSPFDPNRRHCHSPSPDQRLINNNMGFRPKLNFLEINQVK 120
DB	61 QS-GASNDNHVPCYSPFDPNRRHCHSPSPDQRLINNNMGFRPKLNFLEINQVK 119
QY	121 EVTTNGVTTIANLTSTVQVFSDSYQLPYVIGSAHQGCLPPFPADVFMIPQYGLTLN 180
DB	120 EVTNGNTTIANLTSTVQVFTDSYQLPYVIGSAHQGCLPPFPADVFMIPQYGLTLN 179
QY	181 NSGQAVGRSSFCLEYFSPQMLRTGNNFTFSYFEEVPHSSVYASQSLDRLNPLIDQY 240
DB	180 NSGQAVGRSSFCLEYFSPQMLRTGNNFTFSYFEEVPHSSVYASQSLDRLNPLIDQY 239

[illegible]

```

RESULT 2
US-10-038-972A-14
: Sequence 14, Application US/10038972A
: Publication NO. US20020192823A1
: GENERAL INFORMATION:
: APPLICANT: J. Bartlett
: TITLE OF INVENTION: AAV VECTORS AND METHODS
: FILE REFERENCE: 2835/36596US
: CURRENT APPLICATION NUMBER: US/10/038,972A
: CURRENT FILING DATE: 2002-01-04
: PRIOR APPLICATION NUMBER: US 60/260,124
: PRIOR FILING DATE: 2001-01-05
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: SeqId version 3.1
: SEQ ID NO 14
: LENGTH: 598
: TYPE: PRT
: ORGANISM: adeno-associated virus 2 VP2 capsid protein
: US-10-038-972A-14

```

Query Match	85.4%	Score 2481.5	DB 15	Length 598
Best Local Similarity	83.3%	Pred. No. 8.9e-225		
Matches 445, Conservative	37	Mismatches 51	Indels 1	Gaps 1

QY		1	MASGGAPADANNEGADGVGNAGSNHACSTWMLGDEVITTSRTMALPTYNHLKYQIIS	60
Dd		66	MATSGCAPPADNNEGADGVGNASSGNHACSTWMDGXVITTSRTMALPTYNHLKYQIIS	125
QY		61	A\$TGA\$ANDNHVEGYSTPMWGFPDENRPHCHSPRDWRLLNNMWGFAPKRLNFLFLNIQVK	120
Dd		126	QS-GASNDNHVCYSTPMWGFDPENRPHCHSPRDWRLLNNMWGFAPKRLNFLFLNIQVK	184
QY		121	EYTNDGYTTIANNLSTVQVPSDSSEYQLPYLIGSAHOGCLPPRPADVFMIPDYGYTLTN	180
Dd		185	EYTONDGYTTIANNLSTVQVFETDSSEYQLPYVLIGSAHOGCLPPRPADVFMIPDYGYTLTN	244
QY		181	NGSQAVGSSPFCLEYPFPOMLRTGNLFPESTYFEESVPFHSSVAHSOSLDRLMNPILDOY	240
Dd		245	NGSQAVGSSPFCLEYPFPOMLRTGNLFPESTYFEESVPFHSSVAHSOSLDRLMNPILDOY	304
QY		241	LYYLNRTONQSSAONKOLLFSRGSPBAGMSVOPKMLPGCYCQOGRSXTKTIDNNNSNFT	300
Dd		305	LYYLSRTMTPEOSTTYSRLOFQSOAGASDLRDOSRMMLPEPCYRQGRVSXKTSADNNSEYS	364
QY		301	WTGASKYNLNGE\$IIINPGTAMASHKHODEDKCFPM\$GWMI FIKESA\$ASNTALDDNM\$ITD	360
Dd		365	WTGATKYHLNGRSDSLVNPGPAMASHHDEBEKFPO\$GVALIFKQOSEKTVIDEIKYMITD	424
QY		361	EEELATPVPAEREGTYAVNFQSSSTDPA\$TG\$VHAMGLPRGVWODDRVYLOGPIWAKI	420
Dd		425	EEELITTPVALEG\$SV\$TNLORGNROAATADVN\$Q\$VLP\$GVNWODDRDVYLOGPIWAKI	484

[illegible]

```

1  RESULT 3
2  US-10-038-972A-13
3  Sequence 13. Application US/10038972A
4  Publication NO. US20020192823A1
5  GENERAL INFORMATION:
6  APPLICANT: J. Battlet
7  TITLE OF INVENTION: AAV VECTORS AND METHODS
8  FILE REFERENCE: 28335/3696US
9  CURRENT APPLICATION NUMBER: US/10/038,972A
10 CURRENT FILING DATE: 2002-01-04
11 PRIOR APPLICATION NUMBER: 00/260,124
12 PRIOR FILING DATE: 2001-01-05
13 NUMBER OF SEQ ID NOS: 18
14 SOFTWARE: PatentIn version 3.1
15 SEQ ID NO 13
16 LENGTH: 735
17 TYPE: PRT
18 ORGANISM: adeno-associated virus 2 VP1 capsid protein
19 US-10-038-972A-13

```

Query Match	85.4%	Score 2481.5	DB 15	Length 735
Best Local Similarity	83.3%	Pred. No. 1.2e-224		
Matches 445	Conservative 37	Mismatches 51	Indels 1	Gaps 1

QY	1	MAGGAGAPADNNEGADGVAGSAGMWHODSTWJLGRVITSTRWALPTYNNHLYKQISS	60
Db	203	MAGGAPADNNEGADGVAGSAGMWHODSTWJLGRVITSTRWALPTYNNHLYKQISS	262
QY	61	ASTGASDNHRYGYSTPMWGYFDNFNFCHFSBPDWQRLINNMGFRPRKLFKJENIQVK	120
Db	263	OS-GASDNHRYGYSTPMWGYFDNFNFCHFSBPDWQRLINNMGFRPRKLFKJENIQVK	321
QY	121	EVTNNGVTTIANNTLSTVQVPSDEYQLPYLGSAGQCLPPPAVPMIPQYGYTLN	180
Db	322	EVTNNGVTTIANNTLSTVQVPSDEYQLPYLGSAGQCLPPPAVPMIPQYGYTLN	381
QY	181	NGSOAVGRSSFYCLEYFSPQMLRTNNFTFSYFEEVYFHHSSVYHOSLDRLNMPILQY	240
Db	382	NGSOAVGRSSFYCLEYFSPQMLRTNNFTFSYFEEVYFHHSSVYHOSLDRLNMPILQY	441
QY	241	LYYLNFTONQSSAQNKDLFFSRGSPAGMSVQPKWMLPGPCYQOQVSKTKTNDNNSFT	300
Db	442	LYYLSRTNTPSGTTLTQSRLOFQAGASDIRQSRNMLPGPCYQOQVSKTSADNNSNYS	501
QY	301	WTGASKYNLNGEBSIINPGTAMASHKDEDEKFFPKSGYMTIGKSSAGASNTALDNWITD	360
Db	502	WTGATKYHLNGRDSLYNPGPAMASHKDEDEKFFPGQVLLFGKQSEKTIWVDELEKWI	561
QY	361	EEBIRKATNPVABERGTVAVNFQSSSTDPAQGDYAMCALPGMWQORDVYLOGPYAKI	420
Db	562	EEBIRKATNPVABERGTVAVNFQSSSTDPAQGDYAMCALPGMWQORDVYLOGPYAKI	621
QY	421	PHTDGHFHPSPLMGFGGLNPPQLIKNTVPAVNPAPFAFSATKFAFSITQYSTGYVSE	480
Db	622	PHTDGHFHPSPLMGFGGLNPPQLIKNTVPAVNPAPSTTFAAFAPITQYSTGYVSE	681
QY	481	IEWELOKENSKANPBEVQYTSNYAKSANDFTVDNNGLYTEPRIGIRYLTRPL	54
Db	682	IEWELOKENSKANPBEIQYTSNYAKSANDFTVDNNGVYSEPRIGIRYLTRNL	75
RESULT	4		
	US-10-293-478-1		

```
; Sequence 1, Application US/10293478
; Publication No. US2003007841A1
; GENERAL INFORMATION:
; APPLICANT: PATEL, SALIL D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/10/293,478
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/321,589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Adeno-associated virus
; US-10-293-478-1

Query Match      85.4%; Score 2481.5; DB 15; Length 735;
Best Local Similarity 83.3%; Pred. No. 1.2e-224; Indels 1; Gaps 1;
Matches 445; Conservative 37; Mismatches 51;

QY 1 MASGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITTTSTRTWALPTYNNHLYKQISS 60
DB 203 MATGSGAPMADNNEGADGVGNSSGNMHCSTWMDRVTITSTRTWALPTYNNHLYKQISS 262
QY 61 ASFGASNDNHFGYSTWGYFDNRFRHCHSPRDMQRLINNNNGFRPKLNFYKLFNI QVK 120
DB 263 QS-GASNDNHFGYSTWGYFDNRFRHCHSPRDMQRLINNNNGFRPKLNFYKLFNI QVK 321
QY 121 EVTTNDGVTIANNLTSTVOVFSDEYQLPYVLSAQGLCPPPADVFMIPOGYTLN 180
DB 322 EVTQNDGVTIANNLTSTVOVFTDSEYQLPYVLSAQGLCPPPADVFMIPOGYTLN 381
QY 181 NSGOAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEVFFHSSYAHQSGLDRLNMLIDQY 240
DB 382 NSGOAVGRSSFYCLEYFPSSQMLRTGNFTSYTFEEDVPFHSSYAHQSGLDRLNMLIDQY 441
QY 241 LYTLANTONOGSAQONKDLFSGRSPAGMSVOPKRWLPGPCYRQVSKTKTNNNSFT 300
DB 442 LYTLSTRNTPSGTTSQRLQFSQAQASDIDQSHMLPGPCYRQVSKTKSADNNSEYS 501
QY 301 WTSASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVMEFGKESAGASTALDNWITD 360
DB 502 WTSATKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVMEFGKESAGASTALDNWITD 561
QY 361 EEEIKATNPATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWODRDVYLOGPIWAKI 420
DB 562 EEEIRTTNPATQYGSVSTNLQRGNQATADVNTQGVLPGMVWODRDVYLOGPIWAKI 621
QY 421 PHTDGFHSPPLMGFGFLKPNPPOILIKNTVPANPAEPAATKFASTFYQSTGVSV 480
DB 622 PHTDGFHSPPLMGFGFLKPNPPOILIKNTVPANPAEPAATKFASTFYQSTGVSV 681
QY 481 IEMELKENSRRNNPEVQYTSNYAKSANVDPTDNNGLTYNEPPIGRYLTRPL 534
DB 682 IEMELKENSRRNNPEIQTYSNYSKSNVVDPTDNTGVSSEPRPIGRYLTRML 735

RESULT 5
US-10-205-942-4
; Sequence 4, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulek, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205,942
; PRIOR FILING DATE: 2002-07-26
```

```
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Adeno-associated virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2271)
; OTHER INFORMATION: B19/AAV chimeric capsid coding sequence
; US-10-205-942-4

Query Match      16.9%; Score 491.5; DB 15; Length 756;
Best Local Similarity 27.6%; Pred. No. 3e-37; Indels 55; Gaps 16;
Matches 144; Conservative 83; Mismatches 239;

QY 10 ADNNEGADGVGNAS--GNMHCSTWLGDRVITTTSTRTWALPTYNNHLYKQISSASTGASN 67
DB 209 AEAAGAGGGGSSNSVSMWSEGAFTFANSVTCTFSRQFLIPYDPEHHYKVFSPAASCHN 268
QY 68 D-----NHFGYSTWGYFDNRFRHCHSPRDMQRLINNNNGFRPKLNFYKLFNI 117
DB 269 ASGKAQVCTISPMGYSTPWRFLDPAALNLFPSPLEFOHLLENYSIADALTVTITSEI 328
QY 118 QKEAT--TNDGVTIANNLTSTVOVFSDEYQLPYVLSAQGLCPPPADVFMIPOY 175
DB 329 AVVDVTDKTCGGV-QVTDSTTGRCLMLVDHEIKYPPYVLOGGQDTLAPBLPIWYFPPQYA 387
QY 176 YLTLAN-GSQAVG-----RSFYCLEYFPSSQMLRTGNFTSYTFEEVFFHSSYAH 225
DB 388 YLTVDVNTQGISGSKLASBESAFYVLEHSSFOLLGTGTATWGYKFPVPPEMLG 447
QY 226 SGLDLNMLPLIDQYLYLNRKTONQSAQONKDLFSGRSPAGMSVOPKRWLPGPCYRQ 285
DB 448 SQHFEMYNPL---YSGRLGVPDTLGDPKFRSL-----THEDHAIQOPFMGGLVNSV 499
QY 286 RVSKTKTNNNSNFTWTSKYNLNGRESIINPG-TAMASHKDEDEKFFPMGSMVMEFGKE 344
DB 500 STYEGSSNTGAKALTGISTGTSQNTRIISLRGPISQPIHNDTDTKYVTGIAISHQ 559
QY 345 SAGASNTALDNV-----MITDEEIKATNPATERFGTVAVNFQSSSTDPAATGVHAM 397
DB 560 TYG---NAEDKEYQGVGRFPNEKQLQGLNMTY-----FNNKGTQVTDQIE-R 609
QY 398 GALPGMWODRDVYLOGPIWAKIPHTDGFHSP-PLMGFGFLKPNPPOILIKNTVPANP 456
DB 610 PLMWGSVMNRRAALHYSQLSKIPNLDSEFKTQFAPALGQGLHQPPOIFLK--ILFSG 667
QY 457 PAEFSATKFASTFYQSTGVSEIEMEL-QKENSRRNPE 496
DB 668 PIGGISKMGITTYLVQYAVGIMTYMTFKLGPRAKATGRMNP 708

RESULT 6
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12610
LENGTH: 5795
TYPE: PR1
ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 3.7%; Score 108; DB 10; Length 5795;

Best Local Similarity 18.7%; Pred. No. 8.6;
Matches 120; Conservative 85; Mismatches 232; Indels 204; Gaps 34;

Qy 5 GGA PMADNNEGADVGAGNMGHCDSTWLDGRTTSTRWALPTYNHLYKOISSASTG 64
Db 1723 GGA PMADNNEGADVGAGNMGHCDSTWLDGRTTSTRWALPTYNHLYKOISSASTG 64
Qy 65 ASDNHVFGYSTPMGYDPEFRFCH-----FSPRDMORLNNMGRPRKLNFKLNI- 117
Db 1763 -----SVPRHVEETNMGAVFPGVSDYDAKQYKPPNDSTQNAQRMNFQNTSY 1812
Qy 118 -----QVKEVTINDVETIANNLTSTGVFSDSEYQLPYVLSA----- 156
Db 1813 GPSADVIGISTRDRVYDNNQIITKLAKVKEPPRIDNSVTVYAGLTNOQIKNNV 1872
Qy 157 -HOGCLPPFADVYMIPOGYLTNNQSAVGRSSFYCL-EYFPMRLTGNFTFSYTF 214
Db 1873 LSSSSILFPAD-NTP-----LTITNTYSGNNAVTVSALNGVYKARS-----SITM 1922
Qy 215 EEPFSSSYAHSSQSLDLNPLIDQYLYLNRTNOSG-----AKNDLFSRGS- 265
Db 1923 NNVTYTTODEHGRAIDVTRESVD-----SNDATVTVTPQLATTEGAVFIKGD 1973
Qy 266 -----PAGMSV-----QPKMW-----LPBPCY 282
Db 1974 GFDFGVERFTQNPFGATVAMHNDPTWKATGNTKTAIVTLPSCGCTRVNVPVXY 2033
Qy 283 RQGRVSKTKTDNNNSNFTWTGASKYNLNGRESI-----INPGT-----AMASHKDEDK 331
Db 2034 PVANAKAPSRDVKGQNTL-----NGTDAIVYTFDPNTNNGITAAAMANNQCPNNQ 2084
Qy 332 FFPMSGVMITGKESAGSNTALDVMITL--DEEIKATNPVATERGTYAVNFQSSSTD 388
Db 2085 ---OAGVQHLNVDTYTPGITAARVPVTVNYQEPQTSYTTTVG--GTLANGTQ----- 2135
Qy 389 PATGDVYAMGA--LP--GMV--MODRDVYLQGPIMAKI--PHT-----DGHFH 428
Db 2136 -ASGYAMQMANGLPTGFTYKMANNAATGTIDANMAAMANKNAKVYNAKDVLYKHTF 2194
Qy 429 PSLPMGFGGLKN-PPQILIKNTPVPA--NPRAFSATKPSFTIYS-----TGOV 477
Db 2195 ATSLPAFVYVDVQPAKPTVETAAAGITTPGANQVNTHTAGVTTVADLVIKRGNV 2254
Qy 478 SVEIEMELQKNSKRAKPEVOYTSNYAKSANDVTVDNGL 518
Db 2255 ---VTIFTRNNTSPWKE-----ASAATVAGIAGT--NNGI 2286

RESULT 7
US-10-046-583A-1
Sequence 1, Application US/10045583A
Publication No. US20020168743A1

GENERAL INFORMATION:
APPLICANT: Rice, John
APPLICANT: Klotz, Andreas
APPLICANT: Crawford, John
APPLICANT: Lanning, Beth
APPLICANT: Stewart, Sandy
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
TITLE OF INVENTION: SYNTHASE ACTIVITY
FILE REFERENCE: 2037 US Divisional
CURRENT APPLICATION NUMBER: US/10/046,583A
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: 09/626,589
PRIOR FILING DATE: 2000-07-27
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 717
TYPE: PR1
ORGANISM: Arabidopsis thaliana
US-10-046-583A-1

Query Match 3.6%; Score 104.5; DB 15; Length 717;

Best Local Similarity 20.4%; Pred. No. 0.76;
Matches 77; Conservative 54; Mismatches 121; Indels 125; Gaps 18;

Qy 122 VTINDGVTT--IANNLTSTGVFSDSEYQLPYVLSAHGCLPPFADVFMIPQGYLT 178
Db 12 IITKGISTDCKSTSLSSRSLVTD--LP-----SPCLKP----- 45
Qy 179 LNNQSAVGRSSFYCL-----EYF-----PQMLRTGNFTFSYTEEPFSSSYAHSSQ 228
Db 46 -NNNSH-NRBAKVCASLAKEGEYYSNRPTPLDITN-----YIHKNNLSYKE 93
Qy 229 LBLNPLIDQYLYLNRTNOSGSA-----QNKDLFSRSPAGMSV 271
Db 94 LKQSLSELSDVIFNYSKCGHIGSLGVVELTVALHYIFNTPODKILW-----DVGHQS 148
Qy 272 QPKMWLPGEYRQGRVSKTKTDNNNSNFTWTGASKYNLNG--RESINPGTAMASHKOD 328
Db 149 YPKILTG--RKGKPTMKTQNLGSGFTKRGSEHDCCTGHSSTTISAGLMAVGRDL 205
Qy 329 EDFPMSGVMITGKESAGSNTALDVMITDEEIKATN-----PVAT-----ERFG 376
Db 206 KGNNNVVAVIGDAMTAGAYAMANNAGYLDSDMVIINDNKQVSLPTATLGPSPVG 265
Qy 377 TVAVNFQSSSTDAT-----GDVHAMGA-----LPBPCY-----MODRD 409
Db 266 ALSALSRLQSNPALRELBREKAVAKMTQKQIGPWHQLAAKVDYARGMISGTGSLFEELG 325
Qy 410 VYLGPIWAKIPTDGH 426
Db 326 LYITGPV-----DGH 335

RESULT 8
US-09-815-242-12713
Sequence 12713, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haebelbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12713
LENGTH: 2344
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 3.6%; Score 104; DB 10; Length 2344;
Best Local Similarity 20.6%; Pred. No. 5.1;
Matches 77; Conservative 55; Mismatches 154; Indels 88; Gaps 17;

QY 3 SGGGAMADNNEADVGNAAGNMGHCDSTWLGDRVITTSRTTALPTYNHLYKOISSAS 62
DB 387 AGGAGAGAFYTTDSYGA-----SYTSSSTADNAKLNVOPTNN----- 426
QY 63 TGSANDNHYFGYSTPGYDFPNRPHCFSPRDMORLLNNMGFRPRKLNFKLFNIQVKEY 122
DB 427 TPQDPINNG-----DTKMTVKYAGQWTRNI-SWIAKSGTWFSL-----SWT 472
QY 123 TINDGVTTIANNLSTVQVPSDEYQLPVYLSAHOGLPP--PADVEMIPQYGYLTIN 180
DB 473 ASGAGATNLCQVQFGEFEYESAVTVQRYDVTTGKDILPKYSGNVQV-----VTD 527
QY 181 NSGQAGRRSF-----YCLEYFPS-----QMLRTGNNFTFSYTFEEVPHSSVHASOS 228
DB 528 NQGSALTANGVNTSDSSASTYNDNTKTKVKNMGASVTYFTVKAFTVVGQTL 587
QY 229 LRLANPLIDQVLYINRTQNGSAONKDLFSRSGPAGMSV-QPKNMLPGCYR--QQ 285
DB 588 VGTNMPFIV-----LTTDNGTGTYNT-----VTGLPSLSIDSATNSLIGTPYKIGOS 637
QY 286 RVSKTKTD--NNSNFTWTGASKYNLNGRESIIN-----PGTAMASHKDEDEKFFPMG 337
DB 638 TVTVVSTQDQNNKSTTFT-----INVVDTTAFTVTPIGDKSE-VFSPIS 683
QY 338 VMFGESAGASNT 351
DB 684 INIATQDNGNAVT 697

RESULT 9
US-09-815-242-5845
Sequence 5845, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseelbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5845
LENGTH: 1016
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5845

Query Match 3.6%; Score 103.5; DB 10; Length 1016;
Best Local Similarity 21.5%; Pred. No. 1.6;
Matches 122; Conservative 68; Mismatches 196; Indels 181; Gaps 33;

QY 12 NNEGADG-----VGNASGNHCDSTWLGDRVITTSRTTALPT--YNNHLYKOISSAS 62
DB 189 NTTGADGWFPLFSKGAE-EYLTNGGILDKGLVNS-GGFKIDTGYIYSSMDKTEKQAG 246
QY 63 TGSANDNHYFGYSTPGYDFPNRPHCFSPRDMORLLNNMGFRPRKLNFKLFNIQVKEY 122
DB 247 QG-----YRGY-----GAR-----YKNDSSG-----NSQWNGENIDS 274
QY 123 TINDGVTTIANNLSTVQVPSDEYQ-----LPVYLSAHOGLPP--PADVEMIPQYGYLTIN 180
DB 275 KTN--FLVANDSNVT-----SDGKFGRLNDVILTYVASTGKMRLEVAGKTWETSITDL 328
QY 161 -LPPFADVEMIPQYGYLTINNGSQAVSRSSYCLEYFSPQMLRT--GNNFTFS----- 211
DB 329 GSKQAYNPLTTSQGWGLNQGAINANG-----WNRDILKSEFTFTEAPK 375
QY 212 -----YTFEEVPHSSVHASOSLRLANPLIDQVLYINRTQNGS-----SAON-- 256
DB 376 TITTELEKATVEIIPFK-----EKKFNPDLAPGTEKTR-BGQGEKTTITPTLKNPL 426
QY 257 KDLFSRSGP-AGMSVQPKNMLP-----GPCYRQRVSKTKTDNNSNFTWTGASKYN 308
DB 427 TGVITISKGPKEITKDPINELTXYGPETIAPGHDEPDKLP-----GEKEE 475
QY 309 LNGRESIINP--GTAMASHKDEDEKFFPMGCMVIFGKESAGASNTALDNVMTTDEEITA 366
DB 476 VPKKPGIKNPETGDDVVRPVDVTKYKPGKGSIVEKE-----LFFEKERF 523
QY 367 TNPVA--TERFGVANFQSSST-----DPATGDVAMALPGMWQD--RDVYLOGPIM 417
DB 524 NPDLPAGTEKVTREBQGEKTTITPTLKNPLTGVITISKESSEBEITKDPINELTXYGPET 583
QY 418 AKIPHTDGHFHP-----SPLMGFGKLNPPQILIKNTVPVAPNPAFFSATKFSFTT 470
DB 584 ITPGHRD-BFDPKLPFGKEEVPKKGINKNPETGDDVVR-----PPVD-SYTKKQPV-- 632
QY 471 QYSTGQSVIEIWELOKNSKRWNEV 497
DB 633 ---KDSIVEKE-EIPEKERRKFNPD 655

RESULT 10
US-10-198-070-69
Sequence 69, Application US/10198070
Publication No. US20030109437A1
GENERAL INFORMATION:
APPLICANT: AVERBACK, PAUL

APPLICANT: GEMMELL, JACK
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
TITLE OF INVENTION: CELLS
FILE REFERENCE: 59003.000008
CURRENT APPLICATION NUMBER: US/10/198,070
CURRENT FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/306,161
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/306,150
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: 60/331,477
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 69
LENGTH: 1265
TYPE: PRT
ORGANISM: Homo sapiens
US-10-198-070-69

Query Match 3.6%; Score 103.5; DB 15; Length 1265;
Best Local Similarity 19.0%; Pred. No. 2.2;
Matches 114; Conservative 68; Mismatches 180; Indels 237; Gaps 30;

QY 30 STWIG---DVIITSTRTMALPTNNHLYKQI-----SSASTGA 65
DB 49 STFGTLENHKKFYTWEL--HKKHFFENLVSEPLSGSLPALVSQIRLGTTHDTC 106
QY 66 SUNDHYEGYSTPMGYFDFNRFCHSPRDW-----ORLNNWG- 104
DB 107 SEDT---YSTLGRQSEBELRVAEWELECKRIDAVYDEQMTKQRMLETDEMEL 162
QY 105 FRPRKLPKFNIOYKEVTTTNDGYTTIANNLTSTVOVPSDSHYLP-----YLGSAHQ 158
DB 163 FKORRF-----IEBQITNKKAVTGNNETDTMRHVLSSRLSPCCPNCYRRACD 214
QY 159 GCL-----PPFADVF-----MIPOYGL-----TLNNGSQAVGSSSEFCL 194
DB 215 DCSLSHILTCGIMPPVTDDIHHQPLQVDPADYLAERPPVSASSSSSSSPITI 274
QY 195 EYFSPQMLR-TGNNFTSYTFEEV-PHSSYAHSQSLDRMLPLIDOLYYLNTQNSG 252
DB 275 QQHRRLILTSGSAPTFCSDEDDVAPLSAKPAD-----IYPLSNVDTREV 319
QY 253 SAQKDLIFSGSPAGMSVQPKMLPGPCYQOQVSKTKTNNNSNFTWTASKYNLNGR 312
DB 320 VA-NMNGHSELNGGEMWALKDESP-----QISST----- 349
QY 313 ESILNPGTAMASHDDEKFPFPMGCVMIFGKESAGSNTALDNVMTDE--EIKATNP 369
DB 350 -----SSSSSEADDEADESSG-----EPFGAPKE--DVLGSSPTESSKADSP 394
QY 370 VATERFGTVAVNFQSSSTDPATGVDH-----AMGALP--GMWQODRDVYL 412
DB 395 PPS--YPT-----QQAEGAPNTCECHVCKQASGLTPSAMTAYALPCHQGLSPEKPTH- 446
QY 413 QGPIWAKIPTHG--FH-----PSPLMGFGGLKMPPOILIKTTPVAPNPPA----- 458
DB 447 --PALHLYPHIHGVPLHTVPLHPLRPLIHTPLIYATP-----FHSKALPAPVQNTN 498
QY 459 -----EFSATKFASTIQTSGQVSELEMELOKESKRNMP 495
DB 499 KHQVPMASLODHIYPSCRGNTPEMNSKFIPL-----WGSEVMDKMNMP 543

RESULT 11
US-10-108-605-103
Sequence 103, Application US/10108605
Publication No. US20020160934A1
GENERAL INFORMATION:
APPLICANT: Broadus, Julie
APPLICANT: Stam, Lynn

APPLICANT: Bachmann, Jane
TITLE OF INVENTION: KAMDAI, KIM
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
FILE REFERENCE: 31133B
CURRENT APPLICATION NUMBER: US/10/108,605
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: US 09/761,142
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/176,418
PRIOR FILING DATE: 2000-01-14
NUMBER OF SEQ ID NOS: 361
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 103
LENGTH: 3712
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-10-108-605-103

Query Match 3.5%; Score 102; DB 15; Length 3712;
Best Local Similarity 19.1%; Pred. No. 16;
Matches 75; Conservative 55; Mismatches 163; Indels 100; Gaps 16;

QY 103 WGRPRKLNKFL-----FNQYKEVTTNDGYTTIANNLTSTVOVPSDSYQLPYVIGSAH 157
DB 3002 FGGHPRRLNTSISLQPNFDCIDNVVINOQVVDLTVYTG-----GGVE 3045
QY 158 QGCLPPRPADVFNP-QYGYLTLNNGSQAVGSSFCLEVF-----PSOMLRTGNF----- 208
DB 3046 EGGSAAKSTVVSAPHEGYGLRNWVS--SDNNLHVLFHFKTTPRGVIFYAANDQSS 3102
QY 209 TFSYTEEEVPHSSYAHSQSL--DRMLNPLIDOLYYLNTQNO--SSSAQNKDLIFS 262
DB 3103 TIGLSIQDGLLKLNSMSGQVIDRILNDGED-HVTVQHTQGBELRLTVDDVONKRL-- 3158
QY 263 RGSPPAGMSV-----PKMLPGPCYQOQVSKTKTNNNSNFTWTASKY 307
DB 3159 -GSPQPLIEGSDIFPAGLPDNYRTPRMALASLAFVGCISDVTVEEIIINFANSEKKN 3217
QY 308 -NNGRSLNPGTAMASHDDEKFPFPMGCVMIFGKESAGSNTALDNVMTDEEIKI 366
DB 3218 GNING-----CPHVLVLEPSLVPSYSPSGDNEV--EESWNAADTLPLKPDIESTLP 3269
QY 367 TNPATERFGTVAVNFQSSSTDPATGVDHMGALPGMWQODRDVYLQGPWAKIPTHG 426
DB 3270 TTPTTTTTTTTTTTSTTTT-----TTPSPVIVDEKEIEA----- 3310
QY 427 FHSPLMGFGGLKMPPOILIKTTPVAPNPPAE 459
DB 3311 -----KTPQKILITTRPPAKMLPSD 3330

RESULT 12
US-09-836-353A-145
Sequence 145, Application US/09836353A
Publication No. US20030129685A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489PI
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 145
LENGTH: 439
TYPE: PRT

ORGANISM: Homo sapiens
US-09-836-353A-145

Query Match 3.5%; Score 101.5; DB 9; Length 439;

Best Local Similarity 20.9%; Pred. No. 0.69; Mismatches 90; Indels 129; Gaps 20;

Matches 71; Conservative 49; Mismatches 90; Indels 129; Gaps 20;

QY 91 SPDMQRLNNMGFRPRK-----INFKLENIQVKEVTNDGV-----TTIAN 133
DB 53 NPGFTFLDNHMGKQKRLHYSEDELIRLSNITVDH-----EGVYKCYIYSTPFRS 108
QY 134 NLTS-----TVQVPSDSEYQLPYVLSAHQGLPPFPADVFMIPQYGYLTINNGS 183
DB 109 KMTTVEVLAAPSKKVLQVSRDTEGRV--TLSCYQGGKPO-----PQITWL-LNGI 157
QY 184 QAVG-----RSSFCLEYPFSQMLRT-----GNNFTSYTEEVPFH 220
DB 158 QLPEDTRHKLKADGKMTTSTLTVALYGPNSTATCLVHHKALGGKLTPEFQEDVA-- 215
QY 221 SSVAHQSGLDRMLNPLIDQYL--YILNRTQNGSSAQKDLFSGRSPAGMSVQPKMWLP 279
DB 216 RYANNTTPVSTTLE--VDIVSEYVQPTVTTAESDLSNSTDFS----- 256
QY 280 PCYRQGRVSKTKTDNNNSNFT-----WTGASKYNL-NGRESIINPGTAMASHKDEDKF 332
DB 257 PSYPO-----HNGSGATSVAGELSGTSAHHIPEGTEALN-GTV-----TEELF 300
QY 333 -----FPMGVMIFGKESAGASNTALDNWMTDEEIIKA 366
DB 301 RTEASFSEENVTLIS-----IVTFEQDVKS 325

RESULT 13

US-09-984-130-145
Sequence 145, Application US/09984130
Publication No. US20030055231A1
GENERAL INFORMATION:

APPLICANT: Ni et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: P489P2
CURRENT APPLICATION NUMBER: US/09/984,130
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 60/243,792
PRIOR FILING DATE: 2000-10-30
PRIOR APPLICATION NUMBER: 09/836,353
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 149
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 145
LENGTH: 439
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-130-145

Query Match 3.5%; Score 101.5; DB 12; Length 439;

Best Local Similarity 20.9%; Pred. No. 0.69; Mismatches 90; Indels 129; Gaps 20;

Matches 71; Conservative 49; Mismatches 90; Indels 129; Gaps 20;

QY 91 SPDMQRLNNMGFRPRK-----INFKLENIQVKEVTNDGV-----TTIAN 133
DB 53 NPGFTFLDNHMGKQKRLHYSEDELIRLSNITVDH-----EGVYKCYIYSTPFRS 108
QY 134 NLTS-----TVQVPSDSEYQLPYVLSAHQGLPPFPADVFMIPQYGYLTINNGS 183
DB 109 KMTTVEVLAAPSKKVLQVSRDTEGRV--TLSCYQGGKPO-----PQITWL-LNGI 157
QY 184 QAVG-----RSSFCLEYPFSQMLRT-----GNNFTSYTEEVPFH 220

DB 158 QLPEDTRHKLKADGKMTTSTLTVALYGPNSTATCLVHHKALGGKLTPEFQEDVA-- 215
QY 221 SSVAHQSGLDRMLNPLIDQYL--YILNRTQNGSSAQKDLFSGRSPAGMSVQPKMWLP 279
DB 216 RYANNTTPVSTTLE--VDIVSEYVQPTVTTAESDLSNSTDFS----- 256
QY 280 PCYRQGRVSKTKTDNNNSNFT-----WTGASKYNL-NGRESIINPGTAMASHKDEDKF 332
DB 257 PSYPO-----HNGSGATSVAGELSGTSAHHIPEGTEALN-GTV-----TEELF 300
QY 333 -----FPMGVMIFGKESAGASNTALDNWMTDEEIIKA 366
DB 301 RTEASFSEENVTLIS-----IVTFEQDVKS 325

RESULT 14

US-10-080-505-17
Sequence 17, Application US/10080505
Publication No. US20030073166A1
GENERAL INFORMATION:

APPLICANT: St. Gene, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RT/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080,505
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patent In version 3.1
SEQ ID NO 17
LENGTH: 1411
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-080-505-17

Query Match 3.5%; Score 101; DB 15; Length 1411;

Best Local Similarity 18.6%; Pred. No. 4.6; Mismatches 208; Indels 250; Gaps 29;

Matches 123; Conservative 79; Mismatches 208; Indels 250; Gaps 29;

QY 6 GAPMADNN-----EGADGVGNASGNWCHDST-----WTGDRVITSTRWAL 47
DB 521 GATIVNHNATTESTVTITTSDDTINDNTG--DLTKRDIAPFGMGDKDXTKNTGRANV 576
QY 48 PYYNNHLYKQISSASTGASNDHYF-----GYSTPMGYFPENRP 86
DB 577 -TYN-----PLNKDNHFLISGGTNLKGNTDQGTIVFSGRPHAYNNHLNL 623
QY 87 HGFSPDMQRLNNMGFRPRKRLNKLRLNIQVY-----EVTNDGVTTIANLITST 138
DB 624 NELGRPKG-EVVIDDMWIR--TFKAEFQIKGSGTVVSRVSSIEGWTITSNANAT 678
QY 139 VQVPSDSEYQLPYVLSAHQGLPPFPADVFMIPQYGYLTINNGSAGVRSFYCEYFP 198
DB 679 F-----GVVNOQNTICTRSDWTGLTT--CTVNLTDKVIDISIP 716
QY 199 SOMLRTGNNFTSYTEEVPFHSVAHQSGLDRMLN--PLIDQYLYLNRTQNGSSAON 256
DB 717 TTQINGSNLNNAT--VNIH-----GLAKLNGAVTLINHSQTLISNNATQTCNIQL 766
QY 257 KDLIFSRGSPAGMSVQPKMWLPQPCYRQGRVSKTKTDNNNSNFTTWGASKYINNGRESII 316
DB 767 -----SHNANATVDNA--NINGVVHLT 786
QY 317 NPG-----TAMASH--KODEDKFPFPMGVMIFGKESAGASNTALDNWMTDEEIIKA-- 366
DB 787 DSAQFSLKSHFHOQGDQKDTTILENATWTWMPSTTQNLTLNNSYTLNLSAYASASN 846
QY 367 -----TNPVATE-RFGTAVN-----FQSSS-----TDBATPD 393
DB 847 NABRRHSLETETTPYSEEHFNTLVNGLSGQGTFOFTSLIFGYKSDIKILSNDAEGD 906

QY 394 ----VHAMGALPGWMDRDVYLOG-----PTWAKIPHTDGHHPSEPLM---CGF 436
 Db 907 YTLAVRDTGKEP--VTLBQLTLIGLDNOLPDKIKITLKNKHVDAGAMRYELVKNGEFP 964
 QY 437 GLKNPPPOILIKNTVPANPPAEFSATKFSFITYSTGQVSE-----480
 Db 965 RLHNPFIKEGELRNVLVKAQVERALEKQALTTKKQKTEAKVRSKRAAFSDTTPDOSQL 1024
 QY 481 --IEMELOKENSKW-----NPEVQYTSNYA-----KSANVP 511
 Db 1025 NALQAELETTINAOQVAVQAVONOKVTALNQRNEQVKTQDQANLVLATVLEKETAOIDF 1084

RESULT 15
 US-10-098-916A-11
 ; Sequence 11, Application US/10098916A
 ; Publication No. US20030096983A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HECKEL, DAVID G.
 ; APPLICANT: GAHAN, LINDA J.
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A GENE CONFERRING RESISTANCE TO
 ; FILE REFERENCE: CXU-352
 ; CURRENT APPLICATION NUMBER: US/10/098,916A
 ; PRIOR FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 60/276,180
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 1714
 ; TYPE: PRT
 ; ORGANISM: Bombyx mori
 US-10-098-916A-11

Query Match 3.5%; Score 100.5; DB 15; Length 1714;
 Best Local Similarity 19.1%; Pred. No. 6.8;
 Matches 102; Conservative 69; Mismatches 184; Indels 179; Gaps 28;

QY 92 PRDWQRL-----INNNGFRPRKRLNFKLENIQVKEVTTNDGVTTIANNLSTVOVPSDSE 146
 Db 235 PGEMIRMTMTVGIN-----EPLNFEFTNPLHIFSVTALD---SLPNTHTVTLNV----- 279
 QY 147 YQLPYVLGSAHQGLPPPPADVFEMIPQYGLTLN-----NGSQAVGRSSFCLE----- 195
 Db 280 ---QVENEHR---PPRWVEIFAVQOPDEKTAQSPVRAIDGDTGINKPIHYRLTAEE 332
 QY 196 --YFPSQMLRTGNNFTFSY-----TPEEVPFHSS-----YAHSQL----- 229
 Db 333 DTFPHIRKIEGREGSAILIYVPIRDTIQREVFOLSIAYKVDNESSATANVVIYNDI 392
 QY 230 -DRIMNPLIDQY-LYYLNRTQ-----NQGSAQNKDLLFSKSPAGMSVQPKNMLPGFC- 281
 Db 393 NDQRPPEPLFKEYRLINIMETALTINFDQFGRHDDL-----QONAQYVRLIESDPADAA 448
 QY 282 -----YRQQ-----RVGKTKTDN--NNSNFTWTGASK 306
 Db 449 KAFYIAPREVGYQRTFIMGNTANHKMLDYEVEPFRIRLRVIAITDMDNEEHGVAVYVYINL 508
 QY 307 YNLNGRESIINPGTAMASHKDDDEDFPMSGVMI FGKE-SAKASNTALDNVMTDEEIK 365
 Db 509 INMNDDEPIFPHSVQNVSKFTEGKGFVAVNRAHNRDIDDRVHTLMGN----- 558
 QY 366 ATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGWQDR--DYVLQGPINAKIPIHT 423
 Db 559 -----ANNYLSIDKD--TGDH-----VTQDDFFDYHRQSELFVQVRAD 595
 QY 424 D---GHFHPSP---LMGFGGLKNPPQI-LIKNTP-VANPPAEFSATKFSFITYSTG 475
 Db 596 DTLGEPFHTATSOILLIHLEDDINNTPPTLRLPRGSPNVEENVEGYITSEIRATPDPTTA 655
 QY 476 QVSVIEIWELEKENSKRWNPEVQYTSNYAKSAN-VDF-----TVDNNG 517

Db 656 ELRFEIDM-----TTSYATKQGREANPIEPHNCVEIETIYPAINNG 697
 Search completed: July 17, 2003, 18:51:18
 Job time: 27.7143 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 19.1429 Seconds

(without alignments)
2681.720 million cell updates/sec

Title: US-09-807-802a-17

Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRDITGRVLTPL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	74.9%	504	1	VCPV3A coat protein - ade
2	1676	57.7	732	2	S52210 coat protein VP1 -
3	517	17.8	673	1	VCPVB5 coat protein VP1 -
4	487.5	16.8	781	1	VCPV19 coat protein VP1 -
5	234.5	8.1	723	1	VCPVPP coat protein VP1 -
6	232.5	8.0	729	1	VCPVNA coat protein VP1 -
7	232.5	8.0	729	1	A60006 coat protein VP1 -
8	229	7.9	587	1	B44276 coat protein VP1 -
9	226	7.8	722	1	VCPV22 coat protein VP1 -
10	197	6.8	716	1	VCPV2M coat protein VP1 -
11	196.5	6.8	718	1	VCPV1M coat protein VP1 -
12	194	6.7	722	1	VCPVME coat protein VP1 -
13	189	6.7	727	1	VCPV1F coat protein VP1 -
14	189	6.5	584	2	S49594 coat protein VP2
15	189	6.5	727	1	VCPVFP coat protein VP1 -
16	188	6.5	737	1	VCPVCD coat protein VP1 -
17	184	6.3	748	1	VCPVCP coat protein VP1 -
18	177.5	6.1	702	1	VCPVAP coat protein VP1 -
19	175	6.0	722	1	VCPVNC coat protein VP1 -
20	123.5	4.2	690	2	AB0124 probable Tomb-depe
21	122.5	4.2	880	1	F9B6S0 valine-tRNA ligase
22	121	4.2	635	2	F9B660 protein F2X1.10 l
23	118	4.1	739	2	T52289 probable transkeo
24	115.5	4.0	345	1	G97024 probable phospho
25	113.5	3.9	2817	2	B97033 uncharacterized pr
26	109	3.8	418	2	T35753 probable periplasm
27	109	3.8	1072	2	A86827 hypothetical prote
28	109	3.8	1186	2	delta endotoxin -
29	108.5	3.7	1711	2	AB1283 peptidoglycan link

30	108	3.7	655	1	ALKBG cyclomalto-dextrin
31	108	3.7	722	1	A44052 outer layer protei
32	108	3.7	826	1	AD1683 penicillin-binding
33	107.5	3.7	413	2	T23098 hypothetical prote
34	107.5	3.7	648	2	S50856 whn protein - rat
35	107	3.7	777	2	AB0462 probable exported
36	107	3.7	1338	2	T30565 MAP kinase kinase
37	106.5	3.7	403	2	B87513 acyl-CoA dehydroge
38	106	3.6	1271	2	D64237 hypothetical prote
39	105.5	3.6	2500	2	G71609 hypothetical prote
40	105	3.6	1394	2	A29637 position-specific
41	104.5	3.6	559	2	S33724 transcription fact
42	104.5	3.6	717	2	H85171 DEF (C1A1) protein
43	104	3.6	395	2	S50986 MAF1 protein - yea
44	104	3.6	715	2	H90977 hypothetical prote
45	104	3.6	719	2	F85824 hypothetical prote

ALIGNMENTS

RESULT 1

VCPV3A coat protein - adeno-associated virus type 2

C:Species: adeno-associated virus type 2

C:Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jul-1999

C:Accession: A03698

R:Strivastava, A.; Luby, E.W.; Berns, K.I.

U. Virol. 45, 555-564, 1983

A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.

A:Reference number: A03694; MID:83164299; PMID:6300419

A:Accession: A03698

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-504 <SRI>

A:Cross-references: EMBL:J01901; MID:G209616; PID:AAA42376.1; PID:G209621; EMBL:M12405;

A:Superfamily: adeno-associated virus coat protein

C:Keywords: coat protein

Query Match	74.9%	Score 2177.5;	DB 1;	Length 504;
Best Local Similarity	82.5%	Pred. No. 2e-153;		
Matches 392;	Conservative 33;	Mismatches 47;	Indels 3;	Gaps 2;
QY	1	MASGGGAPMADNNEGADGVGASGNHCDSTMGDRVITTSRTPLPTNNHLYKOISS	60	
DB	1	MATGGAHPMADNNEGADGVGASGNHCDSTMGDRVITTSRTPLPTNNHLYKOISS	60	
QY	61	ASTGASNDNHYFGYSTPMGYFDENRPHCHSPRDWQRLINNMGFPKRLNPLFNIOVK	120	
DB	61	QS-GASNDNHYFGYSTPMGYFDENRPHCHSPRDWQRLINNMGFPKRLNPLFNIOVK	119	
QY	121	EYTTNGCVTTIANNLTSTVGVFSDSEYQPLVYVGSAGQCLPPFPADVPIPOYGYLTIN	180	
DB	120	EYVNDGVTIANNLTSTVGVFTDSEYQPLVYVGSAGQCLPPFPADVPIPOYGYLTIN	179	
QY	181	NSGOAVGRSSFFCYLCHFPSCMLRTGNNFPSTFEFVPHSSVAHSOSLDRLNPLIDQY	240	
DB	180	NSGOAVGRSSFFCYLCHFPSCMLRTGNNFTSYFEDVPHSSVAHSOSLDRLNPLIDQY	239	
QY	241	LYYINRTONQSGAONKDLFSRGSPPAGMSVQPKMLPGPCYRQORVSKTKTNNNSFT	300	
DB	240	LYYLSRTNTPSGTTTQSRLOFSGAGASDIRDQSRNMLPGFCYRQORVSKTKSADNNSEYS	299	
QY	301	WTGASKYNLNGRSIINPGTAMASHKDDKDFPMSGWMIFGSGASGASTALDNTWITD	360	
DB	300	WTGATKYNLNGRSLVNP--AMASHKDDKDFPMSGVLIFGSGSGSTKWNIEKWTID	357	
QY	361	EEBKATNPVATRRFGVAVNFQSSSDPATGTVHMGALPGWVWDVYLGPIWAKI	420	
DB	358	EEBIGTNPVATIQYGSVSTNLRGNKQATADVNTQGVLPQWVWDVYLGPIWAKI	417	
QY	421	PHTDGHPSPMLMGFGFLKXNPQIILIKTVPANPAEBSATKFAFTIOYSTG	475	

Db 418 PHTDGHFHSPLMGFGFKHPPOILLIKNTVPANPSTTFSAKFASTITQSTG 472

RESULT 2

552210 coat protein VP1 - muscovy duck parvovirus

N/Alternate names: VPI protein

C/Species: muscovy duck parvovirus

C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-2000

C/Accession: S52210

R/Zadori, Z.; Erdel, J.; Nagy, J.; Kisari, J.

submitted to the EMBL Data Library, September 1993

A/Reference number: S52209

A/Accession: S52210

A/Molecule type: DNA

A/Residues: 1-732 <ZAP>

A/Cross-References: EMBL:X75093; NID:G609091; PIDN:CA52984.1; PID:G609093

A/Experimental source: strain FM

C/Genetics:

A/Gene: VPI

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein

Query Match 57.7%; Score 1676; DB 2; Length 732;
Best Local Similarity 56.9%; Pred. No. 4.8e-116;
Matches 308; Conservative 77; Mismatches 142; Indels 14; Gaps 6;

```
QY 1 MASGGAMADNBEADGVGNASGNWCHDSTWLGDRVITSTRMALPTYNNHLYKQISS 60
DB 199 MABGGSGAMGDSAGAGDGVGNASGNWCHDSTWLGDRVITSTRMALPTYNNHLYKQISS 258
QY 61 ASTGASNDNHFGYSTPMGFDFENRFFCHFSRPMQRLINNNMGPRRLNFKLENIOVK 120
DB 259 GTNPDSN-TQYAGSTPMGFDFENRFFCHFSRPMQRLINNNMGPRRLNFKLENIOVK 317
QY 121 EVTTNDGVTTIANMLSTVQVFSDEYQLPYVLGSAHOGCLPPPADVFMIPOGYGLTLN 180
DB 318 EVTTQDQKTLIANMLSTVQVFSDEYQLPYVLGSAHOGCLPPPADVFMIPOGYGLTLN 377
QY 181 ---NGSAVAGRSFFCLEFYPSOMLRTGNNTFSYTFEEVPHSHASOSLDLMLPLI 237
DB 378 TNOSGARPNDRSAFYCLEFYPSOMLRTGNNTFSYTFEEVPHSHASOSLDLMLPLI 437
QY 238 DQVLYLNRTONQSGAONKOLLFSGSPAGMSVQPKWMLPQPCYRQORVSKTK--TDNN 295
DB 438 DQVLYLNRSEV-NGRNAQ-----FKKAYKAGARGMRMLRFPKLLDORVAYSGGIDNY 491
QY 296 NSNFTWTGASKYKLNIGRESIINPGTAMASHKODEDKFPMSGVMIFGKE--SAGASNTAL 353
DB 492 ANWSIMSKGNKVFLLKDRREYLLQPGVATTTHTEDQASVPAQIITIGIADPYRSGSTLAGI 551
QY 354 DNMMITDEBEIKATNPVATERPGYAVANVFOSSSTDPAGDVHAMGALGQMWQORVDYLO 413
DB 552 SDIMWTDQOEIAPNTNGVWRPVGVLVTNEQNTTAPNAELVIGALGQMWQORNDIYLO 611
QY 414 GPIAKIPIHTGHHFHSPLMGFGFKHPPOILLIKNTVPANPAPSAFASATKFASTITQYS 473
DB 612 GPIAKIPIKTKGKHPSPNLGFGFLANPPPOVFIKNTVPADPPLLEYVNOQMKNSTIYQYS 671
QY 474 TGQVSVLEIEMELQKENSRRANPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGTRYLTRP 533
DB 672 TGQCTVEVWMLERKENSRRANPEIQTFSNGRSTSTWPAWNTGQYVEDRLIGTRYLTQN 731
QY 534 L 534
DB 732 L 732
```

RESULT 3

coat protein VP1 - bovine parvovirus

N/Contains: coat protein VP2

C/Species: bovine parvovirus

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999

C/Accession: A26104

R/Chen, K.C.; Shull, B.C.; Moeses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.

J. Virol. 60, 1085-1097, 1986

A/Title: Complete nucleotide sequence and genome organization of bovine parvovirus.

A/Reference number: A26104; MUID:87061184; PMID:3783814

A/Accession: A26104

A/Molecule type: DNA

A/Residues: 1-673 <CHE>

A/Cross-References: EMBL:M14363; NID:G333454; PIDN:AAB59847.1; PID:G808805

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein

F/118-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 17.8%; Score 517; DB 1; Length 673;
Best Local Similarity 25.1%; Pred. No. 2.9e-30;
Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;

```
QY 5 GGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITSTRMALPTYNNHLYKQISSASTG 64
DB 163 GSGSVGGGGRGGGSGVSTGWTGTFISFENIVTTKNTQRFIDIKNGHLYKS-EVLNTG 221
QY 65 ASNDNHFGYSTPMGFDFENRFFCHFSRPMQRLINNNMGPRRLNFKLENIOKAVTT 124
DB 222 DTAHROY-ALTTPMSYFNFOYSHPSPNDQHLVNDYEFRRKAMIVRYNLIQIKQIMT 280
QY 125 NDGVTTIAN-LTSTVQVFSDEYQLPYVLGSAHOGCLPPPADVFMIPOGYGLTLN 179
DB 281 DGANGVYNNDLTAMGHITPCDGRHRYPYVQHPDDQCMPLPMSIMELPQYAYIAPDISV 340
QY 180 ---NGSAVAGRS-----SFYCLEFYPSOMLRTGNNTFSYTFEEVPHSHASOSLDRL 232
DB 341 VDNNTTIVAEHLLKGVPLWMLNSDHEVLKNG-----R1 375
QY 233 MNPLIDQVLYLNRTON-----QSGAONKOLLFSGSP-----AGMSVQPKWL 277
DB 376 YRIYQLMRMRMRKHQHHIOHASDDVOSTGOKNLLIORTKPNKORFQNALRFSNM 435
QY 278 PGPCYRQORVSKTKTNNNSNFTWTGASKYKLNIGRESIINPGTAMASHKODEDKFPMSG 337
DB 436 SGR-----GIARSTHNAITLOTOSAGALVTMTNGAD---VSG 469
QY 338 VM-----IFGKE-----SAGASNTALDNMMITDEBEIKATNPVAT 372
DB 470 VBAVRGVSTDPILYGGQPSDDLRLRYSASAEGQNPLEN-----AAR 515
QY 373 ERGETAVNFOSSSTDPATGDVHAMGALPGWVQDRDVIQGIYAKIPIHTDGHFHSPL 432
DB 516 HTFTREARTKLIGSNAGDGDYKEMWMLPQOMDSAPISRYNPIWVVPVNRKTLIDQ 575
QY 433 MGGFGLKNPPOILLIK--NTPVPANPAPSAFASATKFASTITQVSGVSEIEMELQKENS 490
DB 576 DGSIPMSHPGCTIFIKLARIPVQNGD-----SFLNIYVTGQVSCVWVEVERGCT 626
QY 491 KRNPEVOYTSNVAKSANVD-FTVNNGLY 519
DB 627 KNRPEYTHS---ATNNSVDAYITINNAGVY 653
```

RESULT 4

VCPV19

coat protein VP1 - parvovirus B19 (strain Au)

C/Species: parvovirus B19

A/Note: host Homo sapiens (man)

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999

C/Accession: A24299

R/Shade, R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.

J. Virol. 58, 921-936, 1986

A/Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated fr

A/Reference number: A24299; MUID:86200451; PMID:3701931

A/Accession: A24299

A/Molecule type: DNA

A/Residues: 1-781 <SHA>


```

QY      3 SGGGAPMADNNEGADVGNASGNMHCDS--TWLGD--RVITTSRTWALPTYNHLYKQ 57
D      174 SGGGGG--GGGGRGAGGVSTGTNNQTEFOYLGELVRIYAHASRLIHLMPHEHYTKR 232
QY      58 IS--SASTGAS--NDNHFGYSTPWCYFDPNRFHCHFSRPMQRLINNMWGRPRKLN 111
D      233 IHLNSESAGVAGVQVODAHQWTPMGLIDANMAGWENPADWQLSNMNTTEINLVSFE 292
QY      112 FKLNIQVKEVT--TNDGVTTIANNTSTVQVPSDSYQLPYVLGSAHQGLPPFPADV 168
D      293 QEIFNVVLKTTTSATSPPTKIYNNDLTASLVALDTNNLTPYPAARSETLGFYPLP 352
QY      169 FMIDQYGL-----TLNNGSAVGRS-----SFYCLE--YFPSQMLRTGNF 208
D      353 TKPQRYRYLSCIRNLNPPYTGOSQOITDSIQGLHSDIMFYTIENAVPIHLRTGDEF 412
QY      209 -TFSYTEEVPFHSSVHSAOSLDR-----LMNPIL--DQYLYLNRTQNGSAGK 257
D      413 STGIYHFDTKPL--KLTHSQWTRNSLGLPKLLEPTTEGDQHPGTLPAANTKGYHQTI 470
QY      258 DILFSRSPAGMSVQPKNMLPGCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESLIN 317
D      471 NNSYTEAT--AIRP-----AQGYNTPYNNFEYSNGGPF-----LT 504
QY      318 P--GTAMASHKDEDEKFFPMGSMIFGKESAGASNTALD--NWTDEEIK--ATNP- 369
D      505 PIVPTADQYNDDEN-----GALRFYMDYGHGLTTSQLEXYTRPQ 549
QY      370 -----VATERGTV--VNFQSS-----TDPATG--DVHMGALP-----GMWQDRDY 410
D      550 SKCGRAPQOQAPLNDENTNNTLLPSDPIGSKSMHFNNTLNTYGPLTALNNTAPV 609
QY      411 YLQSPIAKIHTD--GHFHSPLMGGRGLK--NPPQILIKNTVPANPAPFSA--TKFA 466
D      610 FPNQIMDKELDTLKPRLH--VPAFVCKNNPPGQLFVKIAP--NLTDFFNADSPQ 663
QY      467 SFTQYSTQVSVLEIEMLOKENSKRNPVQYTSNYAKSANVDFTVNNGLYTEPRPIG 526
D      664 PRITTSNFMWKGTLTFTAKRSMNMNPIQOHT-----TANIGNYI--PTNIG 712

```

RESULT 7

coat protein VP1 - porcine parvovirus (strain 90HS)

N:Contains: coat protein VP2

C:Species: porcine parvovirus

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993

C/Accession: A60006

R:Sakurai, M.; Nishimori, T.; Uehimi, C.; Nakajima, H.

Virus Res. 13, 79-86, 1989

A:Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.

A:Reference number: A60006; MUID:89319168; PMID:2750278

A:Accession: A60006

A:Molecule type: DNA

A:Residues: 1-729 <SAK>

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein; glycoprotein

F:151-729/Product: coat protein VP2 #status predicted <VP2>

F:112,198,382,330,433,471,573,604,651/Binding site: carbohydrate (Aen) (covalent) #statu

Query Match 8.0%; Score 232.5; DB 1; Length 729;

Best Local Similarity 23.0%; Pred. No. 3.8e-09;

Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

```

QY      3 SGGGAPMADNNEGADVGNASGNMHCDS--TWLGD--RVITTSRTWALPTYNHLYKQ 57
D      174 SGGGGG--GGGGRGAGGVSTGTNNQTEFOYLGELVRIYAHASRLIHLMPHEHYTKR 232
QY      58 IS--SASTGAS--NDNHFGYSTPWCYFDPNRFHCHFSRPMQRLINNMWGRPRKLN 111
D      233 IHLNSESAGVAGVQVODAHQWTPMGLIDANMAGWENPADWQLSNMNTTEINLVSFE 292
QY      112 FKLNIQVKEVT--TNDGVTTIANNTSTVQVPSDSYQLPYVLGSAHQGLPPFPADV 168

```

```

D      293 QEIFNVVLKTTTSATSPPTKIYNNDLTASLVALDTNNLTPYPAARSETLGFYPLP 352
QY      169 FMIDQYGL-----TLNNGSAVGRS-----SFYCLE--YFPSQMLRTGNF 208
D      353 TKPQRYRYLSCIRNLNPPYTGOSQOITDSIQGLHSDIMFYTIENAVPIHLRTGDEF 412
QY      209 -TFSYTEEVPFHSSVHSAOSLDR-----LMNPIL--DQYLYLNRTQNGSAGK 257
D      413 STGIYHFDTKPL--KLTHSQWTRNSLGLPKLLEPTTEGDQHPGTLPAANTKGYHQTI 470
QY      258 DILFSRSPAGMSVQPKNMLPGCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESLIN 317
D      471 NNSYTEAT--AIRP-----AQGYNTPYNNFEYSNGGPF-----LT 504
QY      318 P--GTAMASHKDEDEKFFPMGSMIFGKESAGASNTALD--NWTDEEIK--ATNP- 369
D      505 PIVPTADQYNDDEN-----GALRFYMDYGHGLTTSQLEXYTRPQ 549
QY      370 VATERGTV--VNFQSS-----TDPATG--DVHMGALP-----GMWQDRDY 410
D      555 APKQOEFMQOAPLNDENTNNTLLPSDPIGSKPMHFNNTLNTYGPLTALNNTAPV 614
QY      416 IWAKIHTD--GHFHSPLMGGRGLK--NPPQILIKNTVPANPAPFSA--TKFASPTQ 471
D      615 IMDKELDTLKPRLH--VPAFVCKNNPPGQLFVKIAP--NLTDFFNADSPQ 668
QY      472 YSTQVSVLEIEMLOKENSKRNPVQYTSNYAKSANVDFTVNNGLYTEPRPIG 526
D      669 YSNFMWKGTLTFTAKRSMNMNPIQOHT-----TANIGNYI--PTNIG 712

```

RESULT 8

coat protein VP1 - parvovirus Luili

C:Species: parvovirus Luili

C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 26-Feb-1999

C/Accession: B44276

R:Diffoot, N.; Chen, K.C.; Bates, R.C.; Lederma, M.

Virology 192, 339-345, 1993

A:Title: The complete nucleotide sequence of parvovirus Luili and localization of a unia

A:Reference number: A44276; MUID:93297126; PMID:8517025

A:Accession: B44276

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-587 <DIF>

A:Cross-References: GB:81888

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein; glycoprotein

F:49190,220,304,371,503,511,514,539,571/Binding site: carbohydrate (Aen) (covalent) #statu

Query Match 7.9%; Score 229; DB 1; Length 587;

Best Local Similarity 22.7%; Pred. No. 5e-09;

Matches 133; Conservative 74; Mismatches 228; Indels 150; Gaps 28;

```

QY      3 SGGGAPMADNNEGADVGNASGNMHCDS--TWLGD--RVITTSRTWALPTYNHLYKQ 52
D      30 SGGG-----GSGGGGVSTGSDYDQTHKFLGDKVETLTASTRVHLMKSEHYCR 83
QY      53 ---HLKQISSASTGASNDNHFGYSTPWCYFDPNRFHCHFSRPMQRLINNMWGRPRK 109
D      84 VRVINTDGTASTAHMAMDHAHQIWM--TPMSLVANMAGWVQPSDWQYISNNMHHINLS 142
QY      110 LNFGLFNIQVKEVT--TNDGVTTIA--NNTLSVQVPSDSYQLPYVLGSAHQGL----- 161
D      143 LDELFNVAIVKTYEQNTGAELKYYNNDLTAMMVALDSNNILPTPAIDNDETFGEY 202
QY      162 --PPFPA-----DVFMIDQYGLTLNNGSAVGRS--SFYCLEYPPS--QMLRTGN 206
D      203 WKPTIPSPYRYVSCDRLNSVYTKDEAGTTIDMGASGINSQFTIENTORINLRTGD 262
QY      207 NF--TFSYTEEVPFHSSVHSAOSLDRIMNPLIDQ-----LYLYLNRTQN- 249

```

```

Db      263 EVATGYVFTPEIRLTHWTQNRHLGQPPQITELPSSDTANATLTARVRSGLTQIGR 322
Qy      250 -----GSGAONKRD-LTFSSGSPAGMSVOPKMLPG-----PCYRQR 286
Db      323 NDVTATRVRAQVGFQCPHDNFTSRAGPKVAVPADITQGLDHDANSLRTYRKQ 382
Qy      287 VSKRTDNNNSNFTWGTASKYNLNGRESIINPTAMASHKDEKFFPMGSMIFGESA 346
Db      383 GQSWASQNNKDRYTM-DAVYD-SGR----- 406
Qy      347 GASTALDNVMTTDEBEIKATNPATER--FGTVAVNPOS--STDPATGDVHMGALP 401
Db      407 WYNNCFIOSVFTSEPN--ANQILTRNDLAGKTDHFTNAFNSYGLTAFPH----- 457
Qy      402 GMWQDRDVLQGGIMAK---IPHTDGHFHSPLMGSGGLKNPPQILIKNTP--VPANP 456
Db      458 -----PAPITQGGIMAKDELDEHKPRHLHTQAPV---CKNNAGQLLVRLAPMLTDQYD 509
Qy      457 PAEFSATKFAFITQYSTGVSVIELEWLOKENSRRNPEVOYTS 501
Db      510 PNSSNLISRIYVYGFPMKGLTKAK---MRPNA-TNNPVPQISA 550

```

RESULT 9

```

VCPV2
coat protein VP1 - parvovirus H1
C:Species: parvovirus H1
A:Note: host Homo sapiens (man)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
C:Accession: A03699
R:Rhode III, S.L.; Paradiso, P.R.
J. Virol. 45, 173-184, 1983
A:Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid
A:Reference number: A03695; MUID:83112183; PMID:6823009
A:Accession: A03699
A:Molecule type: DNA
A:Residues: 1-722 <RHO>
A:Cross-references: EMBL:X01457; EMBL:J02198
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein

```

Query Match 7.8%; Score 226; DB 1; Length 722;

Best Local Similarity 21.7%; Pred. No. 1.1e-08;

Matches 128; Conservative 83; Mismatches 224; Indels 146; Gaps 26;

```

Qy      2 ASGGAPMADNNEGADVGNASGNWCHDSW--IGDRTVITSTRMAL----- 47
Db      153 ADGGGS-SGGGSGGGGIGVSTGYDQTTKPLGDMVEITAHASRLHLGMPSENYC 211
Qy      48 -PTVNNHLYKQISSASTGASNDNHFGYSTPWCYFDFNRFHCHFSRDMQRLINNWGF 105
Db      212 RUVVHNNQTTGHTGKTKVGNNAVYDTHQIW-TFMSLVANAMGWFOPSDMQFIONSESL 270
Qy      106 RPRRLNFKLNIQYKEVTT-----NDGVTTIANNLTSTVOVFSDEYQLPYVLSAHQGC 160
Db      271 NLDSLQGLFNNAVYKVTBQAGQDAIKYNNDLTACMMVALDSNNILPYTPAQSSET 330
Qy      161 L-----PPPPAD--VFMLPQYGYLTLNNGSQ-----AVG-----RSSFYCLE-YFPS 199
Db      331 LGFYPMKPTAPAPRYRYFFMRQLSVTSSNSABCTQITDTIGEQALNSQFFIENTLPT 390
Qy      200 QMLRTGNFTF-SYTFEEVFFHSSYA-----HSQSLRLAN-PLIDQYLYLNRTONQSG 252
Db      391 TLARTGDEFYTYIYFNTDLKLTHWTQNRHLACLOGITDLPTSPDATATSLIANGDRFG 450
Qy      253 SAQNKDLF-----SRGSPAGMSVOPKMLPGPCYRQRVSK 289
Db      451 STQTONVYTEALRTRPAQIGFQPHDNFEARNGGFVVP-----LDI 497
Qy      290 TKTDNNNSNTWGTASKYNLNGR--ESIINPGR-----MASHKDEKFFPMGSG 337
Db      498 TAGEHDHAN---GAIFFNYGKHGHEGMAKQGAAPERIYWDALDSAGRDTARCFV--- 549

```

```

Qy      338 WMIFGESAGASNTALDNVMTDEBEIKA-TNPATERFGTVAVNPOSSTDPATGDVHA 396
Db      550 -----GSAISLIPPNQNLQREDALAGRTNNHNTVFNFSYGLAFHPRDP----- 596
Qy      397 MGALPGMWODRVYLGPIWAK---IPHTDGHFHSPLMGSGGLKNPPQILIKNTPV 453
Db      597 -----IYPGQIWDKELDEHKPRHLHTVAPFV---CKNNPQQLFVHLGP-- 638
Qy      454 ANPAPF--SATKFAFITQYSTGVSVIELEWLOKENSRRNPEVOYTSN 502
Db      639 -NLTDQFPNSTVTSRIVT-YSTFYWKGLIKFKAKLRPLTNPNVYQATTD 687

```

RESULT 10

```

VCPV2M
coat protein VP1 - minute virus of mice
C:Species: minute virus of mice, murine parvovirus
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
C:Accession: A03700
R:Astell, C.R.; Thomson, M.; Merchlinsky, M.; Ward, D.C.
Nucleic Acids Res. 11, 999-1018, 1983
A:Title: The complete DNA sequence of minute virus of mice, an autonomous parvovirus.
A:Reference number: A03696; MUID:83143341; PMID:6298737
A:Accession: A03700
A:Molecule type: DNA
A:Residues: 1-716 <AST>
A:Cross-references: EMBL:V01115
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein

```

Query Match 6.8%; Score 197; DB 1; Length 716;

Best Local Similarity 20.0%; Pred. No. 1.6e-06;

Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;

```

Qy      3 SGGAPMADNNEGADVGNASGNWCHDS--TWLGD---RVITSTRMALPTVNNHLYKQ 57
Db      160 SGGG-----GSGGGGAGVSTGSDYDQTHYRFLGDMVEITATLRLVHLNMPKSENYC 213
Qy      58 I-----SSASTGASNDNHFGYSTPWCYFDFNRFHCHFSRDMQRLINNWGFPRKRL 110
Db      214 IRVHNTTDSVKNMMAKDAHEQIWTWPSLVANAMGWMLQPSDMQYICNTMSQLNLVSL 273
Qy      111 NPLFNIOYKEVTTND-----GVTTIANNLTSTVOVFSDEYQLPYVLSAHQGCLEPPFA 166
Db      274 DQEIFNVVLKTYTEQDLOGQALIKYNNDLTACMMVALDSNNILPYTPAQSMTGLGFYFW 333
Qy      167 DVEMIPQYQY-----LTLNNGSQAV-----GRSSFYCLEYFPS-QMLRTGN 207
Db      334 KPTIASPYRYFVDRDLSTVYENQEGTYEHNNWGTGKGPQGFIENTQOITLARTGE 393
Qy      208 F-TFSTYTFEEVFFHSSYAHSQSLRLAN-PLIDQYLYLNRTONQSGAONKDLFSSRG 264
Db      394 FATGYFYFDTSV--KLTHWTQNRQLGQRPILSTF---PEADTAGT-----LTAQG 441
Qy      265 SPAGMSVOPKML-----PGCYRQRVSKTKTDNNNSNFT 300
Db      442 SRHTQTMGVNVMSEAIRTRPAQVGFQPHDNFEARNGF-FAAPVPADITQGVDEAN 500
Qy      301 WTGASKYNLNGRESIINPG-----TAMASHKDEKDFP-----PMSGVMTF 341
Db      501 GSVRYSYGKHGHEGMAHSHGAPARYTWDETSFSSGSDTQDGFQSAPLVVPPLNGI--- 557
Qy      342 GKESAGASNTALDNVMTDEBEIKATNPATERFGTVAVNPOS--STDPATGDVHMGGA 399
Db      558 -----LTNANPIGTKN---DIHFSVFNFSYGLTAFSH--- 587
Qy      400 LPMWQDRDVLQGGIMAK---IPHTDGHFHSPLMGSGGLKNPPQILIKNTP--VPA 454
Db      588 -----PSPVTFQGGIMAKDELDEHKPRHLHTAPV---CKNNAGQLLVRLAPMLTQ 637
Qy      455 NPAEFSATKFAFITQYSTGVSVIELEWLOKENSRRNPEVOYTSNKAASANDFTVD 514
Db      638 YDENGATLSRIYVYGFPMKGLTWAKLRA---NTNPNPVQYVAB-----D 682

```


coat protein VP1 - feline panleukopenia virus
N:Contains: coat protein VP2

C:Species: feline panleukopenia virus, FPLV

C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999

C:Accession: A03701

R:Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winston, S.; Hahn, W.

J. Virol. 55, 574-587, 1985

A:Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parv

A:Reference number: A03697; MUID:85265017; PMID:2991581

A:Accession: A03701

A:Molecule type: DNA

A:Residues: 1-727 <CAR>

A:Cross-references: EMBL:M10824; NID:G333474; PID:AAA47161.1; PID:G333476

C:Genetics:

A:Intons: 11/1

C:Superfamily: parvovirus coat protein

C:Keywords: coat protein

F:14-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.7%; Score 194; DB 1; Length 727;

Best Local Similarity 20.7%; Pred. No. 2.7e-06;

Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;

```

QY 5 GGAPMADNNEGADGVNAGSNMHCSTWLGDRVITST-----RTMALPTVNN- 52
DB 153 GGQP-AVRNERATSGSGGGGGGS--GGVGISTGTENNQTEFKLENGWVEITANS 209
QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPGYFDFNRFHCHGPRDW 95
DB 210 RLVLHNMPESENKRRVAVNNMDKTAIVGNMALLDTHAQIVTWSLVANAMGVWFNPGDW 269
QY 96 QRLINNNMGFRPKRLNFKLFIQVEVT--TNDGVTIANLSTVQVSDSEYOLPYV 152
DB 270 QLVNTMSEHLVSPFOEIRNVVLKTVSESATQPPTKVYNNDLTSLMALDSDNNTMFT 329
QY 153 LGSAGHGCCL-----PPPPA-----DVFMIPQYGYLTLLNGSOAVGSS----- 189
DB 330 PAMRSEBTGLFYPWKPTIIPWRYFYQMDRTLIPSH-----TGTSTPTNIVHGTDPPD 363
QY 190 -SFYCLE-YFPGOMLTGNNF--TF-----STFEE-----VPHSSYASHOSL 229
DB 384 VQFTYTIENSVPVHLRLTGDEFATGTFPDCPKRLTHWTQNRALGLPPLNLSLPSGGA 443
QY 230 DR-----LNNPLIDQYL--YYLNRTONOS----- 251
DB 444 TNGGDIGVQDQKRGVTOGNTDITTEATIRPAEYGSAPYISFEASTGPRKIPITAG 503
QY 252 -GSAQNKDLLFSRGSFAGMSVQPRNMLPGCY--RQQRVSKTKTDNNNSNFTWTSKSY 307
DB 504 RGAQGTDEMOAADDP--VAAGRQHGQKTTTGTETPERFTY----- 543
QY 308 NUNGRESIINPGTAMASHKDEDEKFFPMGSGVMTFGKSAGASNTALD-NVMTIDEERIK 366
DB 544 -----IAHQDT-----GRYPADWTQININPNIPTVNDVNLTP 575
QY 367 TNVATERFETVAVNFQSSSTDPATGDUHMGALPGWVQDQDLYLQGPIMAKIPHTDGH 426
DB 576 TDIIG--GKTGINY--TNIPTGYPLTALNVP-----PYTPNGQIMDKSPDID-- 620
QY 427 FHSPLMGFGFLK-----NPPQILIKTNPVPA--PPAEFSATKFAFIT 470
DB 621 -----LKPRLHVNAPFVCQNNCPQGLFVKVAPNLTNEVDPAASNMGR-----IV 665
QY 471 QYSTGVSVIEIEMELQENSKRNPEVQYTSNKAASANDFTYDNNGLYTERPPIG 526
DB 666 TYSDFWPKGLVFKAKLRASHATNPPIQOMSIN-----VDNQFNIV-PNNIG 710

```

RESULT 14

S49594

capsid protein VP2 - canine parvovirus

C:Species: canine parvovirus, CPV

C:Date: 05-Mar-1995 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999

C:Accession: S49594

R:Mizak, B.; Plucieniczak, A.

submitted to the EMBL Data Library, November 1994

A:Description: Polish isolates of canine parvovirus.

A:Reference number: S49594

A:Accession: S49594

A:Molecule type: DNA

A:Residues: 1-584 <MIZ>

A:Cross-references: EMBL:Z46651; NID:G572660; PID:CAA86612.1; PID:G572661

C:Superfamily: parvovirus coat protein

Query Match 6.5%; Score 189; DB 2; Length 584;

Best Local Similarity 20.8%; Pred. No. 4.5e-06;

Matches 132; Conservative 89; Mismatches 223; Indels 192; Gaps 33;

```

QY 5 GGAPMADNNEGADGVNAGSNMHCSTWLGDRVITST-----RTMALPTVNN- 52
DB 10 GGQP-AVRNERATSGSGGGGGGS--GGVGISTGTENNQTEFKLENGWVEITANS 66
QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPGYFDFNRFHCHGPRDW 95
DB 67 RLVLHNMPESENKRRVAVNNMDKTAIVGNMALLDTHAQIVTWSLVANAMGVWFNPGDW 126
QY 96 QRLINNNMGFRPKRLNFKLFIQVEVT--TNDGVTIANLSTVQVSDSEYOLPYV 152
DB 127 QLVNTMSEHLVSPFOEIRNVVLKTVSESATQPPTKVYNNDLTSLMALDSDNNTMFT 186
QY 153 LGSAGHGCCL-----PPPPA-----DVFMIPQYGYLTLLNGSOAVGSS----- 189
DB 187 PAMRSEBTGLFYPWKPTIIPWRYFYQMDRTLIPSH-----TGTSTPTNIVHGTDPPD 240
QY 190 -SFYCLE-YFPGOMLTGNNF--TFSTFEEVPHSSYASHOSLRLNPLIDQYLTYLR 246
DB 241 VQFTYTIENSVPVHLRLTGDEFATGTFPDCPKRLTHWTQNRALGLPPLNLSLPSQ 296
QY 247 TQNGS-----GSAQNK-----DLFSRGSFAGMSVQPRNMLPGCYRQQRV 287
DB 297 SEBGTNPGYIGVQDQKRGVTOGNTDITTEATIRPAEYGSAPYISFEASTGPRKIPITAG 348
QY 288 SK-----TKTDNNNSNFTWTSKSY--NLNGRESIINPGT--AMASHKD 327
DB 349 TQGPFPKPIAAGRGAGQTDENQAA--DSDPRAFPGRQHGQKTTTGTETPERFTYIAHOD 405
QY 328 DEDKFFPMGSGVMTFGKESAGASNTALD-NVMTIDEERIKATNPVATERFETVAVNFQSS 386
DB 406 T-----GRYPADWTQININPNIPTVNDVNLTPDITG--GKTGINY--TN 446
QY 387 TDPATGDVHMGALPGWVQDQDLYLQGPIMAKIPHTDGHFHSPLMGFGFLK----- 439
DB 447 IFNTYGPLTALNVP-----PYTPNGQIMDKSPDID-----LKPRLHVN 486
QY 440 -----NPPQILIKTNPVPA--PPAEFSATKFAFITQYSTGVSVIEIEMELQENS 490
DB 487 PFVCQNNCPQGLFVKVAPNLTNEVDPAASNMGR-----IVTYSDFWPKGLVFKAKLRAS 542
QY 491 KRNPEVQYTSNKAASANDFTYDNNGLYTERPPIG 526
DB 543 HTWNPPIQOMSIN-----VDNQFNIV-PSNIG 567

```

RESULT 15

VCPVFP

coat protein VP1 - feline panleukopenia virus (strain 193)

N:Contains: coat protein VP2

C:Species: feline panleukopenia virus, FPLV

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 12-Apr-1996

C:Accession: B36608

R:Marlyn, J.C.; Davidson, B.E.; Studdert, M.J.

J. Gen. Virol. 71, 2747-2753, 1990

A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvo

A:Reference number: A36608; MUID:91073139; PMID:2174965

A:Accession: B36608

A:Molecule type: DNA

A:Residues: 1-727 <MAR>
A/Cross-references: GB:X55115
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein
F:144-727/Product: coat protein vp2 #status predicted <vp2>

Query Match 6.5%; Score 189; DB 1; Length 727;
Best Local Similarity 20.6%; Pred. No. 6.3e-06;
Matches 135; Conservative 73; Mismatches 216; Indels 232; Gaps 32;

```
QY 5 GGAPMADNNEGADGVNAGNMGHCDSTWLGDRVITST-----RTMALPTVNN- 52
Db 153 GGQP-AVNERATSGGNGGGGGS--GGVGISTGFNNQTEKFLNGMVEITANSS 209
QY 53 ---HL-----YQI-----SSASTGASNDNHFGYSTPNCGYFPDNRHCHSPRDW 95
Db 210 RLVLNMPESBENYKRVVNNMDKTAIVKGNALDDIHQIVTPMSLVANAMGVFNPGDW 269
QY 96 ORLINNMGFRPKLNFKNIOVKEVT---TNDGVTTIANNLTSTVOFSDSEYOLPYV 152
Db 270 QLVNTMSELHLVSEFQEIFNVVLKTVSESATQPTKYNNDLTASLMLALDSNNMTMFT 329
QY 153 LGSNAHQCGL-----PPPPA-----DVFMIPOGYLTLLNGSOAVGRS----- 189
Db 330 PAAMRSFTLGFPWKPTIPFPMRYFPQMDRTLPSH-----TGTSGPTNVYHGTDPDD 383
QY 190 -SPYCLE-YFPSCMLRTGNNF---TF-----SYTFEE-----VPPHSSVAHSOSL 229
Db 384 VQFTTIENSVAHLRTGDEPATGTFFDCKPCRLTHWQTNRALGLPPLNLSLPOSEGA 443
QY 230 DR-----LMNPLIDQYL--YILNRTQNS----- 251
Db 444 TNFGDIGVQDKRGRVYQMGNTDYITTEATIRPAVGVSAFYSGFEASTQGPFTPIAAG 503
QY 252 -GSAQNKDLLFSRGSPPAGMSVQPPNMLPGCY--RQQRVSKYKTDNNNSNFTWTGASKY 307
Db 504 RGGAQTDENQADDDPR-----YAFGRHQGQKTTTGTGETPERFTY----- 543
QY 308 NINRGESITNGTMAHSHKDEDEKFFPMISGVMIFGKESAGASNTALD-NVMITDEERIKA 366
Db 544 -----IAHQDT-----GRYPEGDIQININFNLEVINDNVLLP 575
QY 367 TNPVATERFGTVAVNFOSSSTDPAQDVHAMGALPGMWQDRDVLQSPITAKIPIHTDGH 426
Db 576 TDPETG---GKTGINY--TNIFNTYGPULTANNP-----PYYPNGQIMDKKEFDTD-- 620
QY 427 FHPSPILMGFGGLK-----NPPQILIKNTPPVPAK---PPAFESATKPAASFTT 470
Db 621 -----LKPRLHVNAAPFVCCNNCPGQLFVKVAPNLITNEYDPDASANMSR-----IV 665
QY 471 QYSTQGVSEIEIEMLOKENSRRKNPEVOYTSNVAKSANVDFTVDDNGLYTEPRPIG 526
Db 666 TYSDFMWMKGLVFAKLRASHHTNPDIQMSIN-----VDNQFNVTY-PNNIG 710
```

Search completed: July 17, 2003, 18:35:48
Job time : 21.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 7011 Seconds
(without alignments)
10896.630 Million cell updates/sec

Title: US-09-807-802a-1
Perfect score: 4718
Sequence: 1 ttcgccacccctctctgcg.....cgcagagagagtg99gcaaa 4718

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estdb: *
2: em_estdb: *
3: em_estdb: *
4: em_estdb: *
5: em_estdb: *
6: em_estdb: *
7: em_estdb: *
8: em_estdb: *
9: gb_est1: *
10: gb_est2: *
11: gb_est3: *
12: gb_est4: *
13: gb_est5: *
14: gb_est6: *
15: em_estfun: *
16: em_estcom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.8	1.9	489	17	AZ289042
2	76.2	1.6	753	17	BH115587
3	55.6	1.2	493	17	AZ028362
4	55.2	1.2	636	13	BM440166
5	55.2	1.2	659	13	BM371945
6	55.2	1.2	692	14	BQ766288

7	55.2	1.2	852	12	BG310300	BG310300	HVSMC001
8 <td>54<td>1.1<td>652<td>13<th>Bj268130</th><th>Bj268130</th><th></th></td></td></td></td>	54 <td>1.1<td>652<td>13<th>Bj268130</th><th>Bj268130</th><th></th></td></td></td>	1.1 <td>652<td>13<th>Bj268130</th><th>Bj268130</th><th></th></td></td>	652 <td>13<th>Bj268130</th><th>Bj268130</th><th></th></td>	13 <th>Bj268130</th> <th>Bj268130</th> <th></th>	Bj268130	Bj268130	
9 <td>53.2<td>1.1<td>607<td>14<th>B0660503</th><th>B0660503</th><th>H1020054</th></td></td></td></td>	53.2 <td>1.1<td>607<td>14<th>B0660503</th><th>B0660503</th><th>H1020054</th></td></td></td>	1.1 <td>607<td>14<th>B0660503</th><th>B0660503</th><th>H1020054</th></td></td>	607 <td>14<th>B0660503</th><th>B0660503</th><th>H1020054</th></td>	14 <th>B0660503</th> <th>B0660503</th> <th>H1020054</th>	B0660503	B0660503	H1020054
10 <td>52.4<td>1.1<td>528<td>9<th>AL818251</th><th>AL818251</th><th></th></td></td></td></td>	52.4 <td>1.1<td>528<td>9<th>AL818251</th><th>AL818251</th><th></th></td></td></td>	1.1 <td>528<td>9<th>AL818251</th><th>AL818251</th><th></th></td></td>	528 <td>9<th>AL818251</th><th>AL818251</th><th></th></td>	9 <th>AL818251</th> <th>AL818251</th> <th></th>	AL818251	AL818251	
11 <td>52.4<td>1.1<td>688<td>14<th>BM616637</th><th>BM616637</th><th>HB02E06_T</th></td></td></td></td>	52.4 <td>1.1<td>688<td>14<th>BM616637</th><th>BM616637</th><th>HB02E06_T</th></td></td></td>	1.1 <td>688<td>14<th>BM616637</th><th>BM616637</th><th>HB02E06_T</th></td></td>	688 <td>14<th>BM616637</th><th>BM616637</th><th>HB02E06_T</th></td>	14 <th>BM616637</th> <th>BM616637</th> <th>HB02E06_T</th>	BM616637	BM616637	HB02E06_T
12 <td>52.4<td>1.1<td>1065<td>11<th>AY110092</th><th>AY110092</th><th>Zea mays</th></td></td></td></td>	52.4 <td>1.1<td>1065<td>11<th>AY110092</th><th>AY110092</th><th>Zea mays</th></td></td></td>	1.1 <td>1065<td>11<th>AY110092</th><th>AY110092</th><th>Zea mays</th></td></td>	1065 <td>11<th>AY110092</th><th>AY110092</th><th>Zea mays</th></td>	11 <th>AY110092</th> <th>AY110092</th> <th>Zea mays</th>	AY110092	AY110092	Zea mays
13 <td>52<td>1.1<td>877<td>13<th>BI952167</th><th>BI952167</th><th>HVSMC000</th></td></td></td></td>	52 <td>1.1<td>877<td>13<th>BI952167</th><th>BI952167</th><th>HVSMC000</th></td></td></td>	1.1 <td>877<td>13<th>BI952167</th><th>BI952167</th><th>HVSMC000</th></td></td>	877 <td>13<th>BI952167</th><th>BI952167</th><th>HVSMC000</th></td>	13 <th>BI952167</th> <th>BI952167</th> <th>HVSMC000</th>	BI952167	BI952167	HVSMC000
14 <td>51.6<td>1.1<td>570<td>10<th>BE452865</th><th>BE452865</th><th>894066D03</th></td></td></td></td>	51.6 <td>1.1<td>570<td>10<th>BE452865</th><th>BE452865</th><th>894066D03</th></td></td></td>	1.1 <td>570<td>10<th>BE452865</th><th>BE452865</th><th>894066D03</th></td></td>	570 <td>10<th>BE452865</th><th>BE452865</th><th>894066D03</th></td>	10 <th>BE452865</th> <th>BE452865</th> <th>894066D03</th>	BE452865	BE452865	894066D03
15 <td>51.6<td>1.1<td>612<td>13<th>BI873856</th><th>BI873856</th><th>963111F11</th></td></td></td></td>	51.6 <td>1.1<td>612<td>13<th>BI873856</th><th>BI873856</th><th>963111F11</th></td></td></td>	1.1 <td>612<td>13<th>BI873856</th><th>BI873856</th><th>963111F11</th></td></td>	612 <td>13<th>BI873856</th><th>BI873856</th><th>963111F11</th></td>	13 <th>BI873856</th> <th>BI873856</th> <th>963111F11</th>	BI873856	BI873856	963111F11
16 <td>51.6<td>1.1<td>655<td>13<th>BI875610</th><th>BI875610</th><th>963125H10</th></td></td></td></td>	51.6 <td>1.1<td>655<td>13<th>BI875610</th><th>BI875610</th><th>963125H10</th></td></td></td>	1.1 <td>655<td>13<th>BI875610</th><th>BI875610</th><th>963125H10</th></td></td>	655 <td>13<th>BI875610</th><th>BI875610</th><th>963125H10</th></td>	13 <th>BI875610</th> <th>BI875610</th> <th>963125H10</th>	BI875610	BI875610	963125H10
17 <td>51.6<td>1.1<td>680<td>12<th>BG844693</th><th>BG844693</th><th>1024007D1</th></td></td></td></td>	51.6 <td>1.1<td>680<td>12<th>BG844693</th><th>BG844693</th><th>1024007D1</th></td></td></td>	1.1 <td>680<td>12<th>BG844693</th><th>BG844693</th><th>1024007D1</th></td></td>	680 <td>12<th>BG844693</th><th>BG844693</th><th>1024007D1</th></td>	12 <th>BG844693</th> <th>BG844693</th> <th>1024007D1</th>	BG844693	BG844693	1024007D1
18 <td>51<td>1.1<td>536<td>10<th>BE603252</th><th>BE603252</th><th>HVSMC010</th></td></td></td></td>	51 <td>1.1<td>536<td>10<th>BE603252</th><th>BE603252</th><th>HVSMC010</th></td></td></td>	1.1 <td>536<td>10<th>BE603252</th><th>BE603252</th><th>HVSMC010</th></td></td>	536 <td>10<th>BE603252</th><th>BE603252</th><th>HVSMC010</th></td>	10 <th>BE603252</th> <th>BE603252</th> <th>HVSMC010</th>	BE603252	BE603252	HVSMC010
19 <td>50.6<td>1.1<td>500<td>12<th>BG905125</th><th>BG905125</th><th>Talx1137D</th></td></td></td></td>	50.6 <td>1.1<td>500<td>12<th>BG905125</th><th>BG905125</th><th>Talx1137D</th></td></td></td>	1.1 <td>500<td>12<th>BG905125</th><th>BG905125</th><th>Talx1137D</th></td></td>	500 <td>12<th>BG905125</th><th>BG905125</th><th>Talx1137D</th></td>	12 <th>BG905125</th> <th>BG905125</th> <th>Talx1137D</th>	BG905125	BG905125	Talx1137D
20 <td>49.4<td>1.0<td>619<td>14<th>BQ822998</th><th>BQ822998</th><th>103105D0</th></td></td></td></td>	49.4 <td>1.0<td>619<td>14<th>BQ822998</th><th>BQ822998</th><th>103105D0</th></td></td></td>	1.0 <td>619<td>14<th>BQ822998</th><th>BQ822998</th><th>103105D0</th></td></td>	619 <td>14<th>BQ822998</th><th>BQ822998</th><th>103105D0</th></td>	14 <th>BQ822998</th> <th>BQ822998</th> <th>103105D0</th>	BQ822998	BQ822998	103105D0
21 <td>49.4<td>1.0<td>665<td>13<th>BI966796</th><th>BI966796</th><th>103104H1</th></td></td></td></td>	49.4 <td>1.0<td>665<td>13<th>BI966796</th><th>BI966796</th><th>103104H1</th></td></td></td>	1.0 <td>665<td>13<th>BI966796</th><th>BI966796</th><th>103104H1</th></td></td>	665 <td>13<th>BI966796</th><th>BI966796</th><th>103104H1</th></td>	13 <th>BI966796</th> <th>BI966796</th> <th>103104H1</th>	BI966796	BI966796	103104H1
22 <td>49.4<td>1.0<td>787<td>11<th>AY104424</th><th>AY104424</th><th>Zea mays</th></td></td></td></td>	49.4 <td>1.0<td>787<td>11<th>AY104424</th><th>AY104424</th><th>Zea mays</th></td></td></td>	1.0 <td>787<td>11<th>AY104424</th><th>AY104424</th><th>Zea mays</th></td></td>	787 <td>11<th>AY104424</th><th>AY104424</th><th>Zea mays</th></td>	11 <th>AY104424</th> <th>AY104424</th> <th>Zea mays</th>	AY104424	AY104424	Zea mays
23 <td>49<td>1.0<td>485<td>12<th>BF728378</th><th>BF728378</th><th>1000061E1</th></td></td></td></td>	49 <td>1.0<td>485<td>12<th>BF728378</th><th>BF728378</th><th>1000061E1</th></td></td></td>	1.0 <td>485<td>12<th>BF728378</th><th>BF728378</th><th>1000061E1</th></td></td>	485 <td>12<th>BF728378</th><th>BF728378</th><th>1000061E1</th></td>	12 <th>BF728378</th> <th>BF728378</th> <th>1000061E1</th>	BF728378	BF728378	1000061E1
24 <td>49<td>1.0<td>637<td>10<th>AM399968</th><th>AM399968</th><th>707055F01</th></td></td></td></td>	49 <td>1.0<td>637<td>10<th>AM399968</th><th>AM399968</th><th>707055F01</th></td></td></td>	1.0 <td>637<td>10<th>AM399968</th><th>AM399968</th><th>707055F01</th></td></td>	637 <td>10<th>AM399968</th><th>AM399968</th><th>707055F01</th></td>	10 <th>AM399968</th> <th>AM399968</th> <th>707055F01</th>	AM399968	AM399968	707055F01
25 <td>49<td>1.0<td>686<td>13<th>BM610226</th><th>BM610226</th><th>170006871</th></td></td></td></td>	49 <td>1.0<td>686<td>13<th>BM610226</th><th>BM610226</th><th>170006871</th></td></td></td>	1.0 <td>686<td>13<th>BM610226</th><th>BM610226</th><th>170006871</th></td></td>	686 <td>13<th>BM610226</th><th>BM610226</th><th>170006871</th></td>	13 <th>BM610226</th> <th>BM610226</th> <th>170006871</th>	BM610226	BM610226	170006871
26 <td>49<td>1.0<td>599<td>12<th>BE726188</th><th>BE726188</th><th>894089F04</th></td></td></td></td>	49 <td>1.0<td>599<td>12<th>BE726188</th><th>BE726188</th><th>894089F04</th></td></td></td>	1.0 <td>599<td>12<th>BE726188</th><th>BE726188</th><th>894089F04</th></td></td>	599 <td>12<th>BE726188</th><th>BE726188</th><th>894089F04</th></td>	12 <th>BE726188</th> <th>BE726188</th> <th>894089F04</th>	BE726188	BE726188	894089F04
27 <td>48.8<td>1.0<td>672<td>12<th>BG859092</th><th>BG859092</th><th>1024061C1</th></td></td></td></td>	48.8 <td>1.0<td>672<td>12<th>BG859092</th><th>BG859092</th><th>1024061C1</th></td></td></td>	1.0 <td>672<td>12<th>BG859092</th><th>BG859092</th><th>1024061C1</th></td></td>	672 <td>12<th>BG859092</th><th>BG859092</th><th>1024061C1</th></td>	12 <th>BG859092</th> <th>BG859092</th> <th>1024061C1</th>	BG859092	BG859092	1024061C1
28 <td>48.8<td>1.0<td>1137<td>13<th>BI952212</th><th>BI952212</th><th>HVSMC000</th></td></td></td></td>	48.8 <td>1.0<td>1137<td>13<th>BI952212</th><th>BI952212</th><th>HVSMC000</th></td></td></td>	1.0 <td>1137<td>13<th>BI952212</th><th>BI952212</th><th>HVSMC000</th></td></td>	1137 <td>13<th>BI952212</th><th>BI952212</th><th>HVSMC000</th></td>	13 <th>BI952212</th> <th>BI952212</th> <th>HVSMC000</th>	BI952212	BI952212	HVSMC000
29 <td>48.6<td>1.0<td>495<td>12<th>BF630386</th><th>BF630386</th><th>HVSMC000</th></td></td></td></td>	48.6 <td>1.0<td>495<td>12<th>BF630386</th><th>BF630386</th><th>HVSMC000</th></td></td></td>	1.0 <td>495<td>12<th>BF630386</th><th>BF630386</th><th>HVSMC000</th></td></td>	495 <td>12<th>BF630386</th><th>BF630386</th><th>HVSMC000</th></td>	12 <th>BF630386</th> <th>BF630386</th> <th>HVSMC000</th>	BF630386	BF630386	HVSMC000
30 <td>48.2<td>1.0<td>592<td>12<th>BE776688</th><th>BE776688</th><th>MY-19-B-0</th></td></td></td></td>	48.2 <td>1.0<td>592<td>12<th>BE776688</th><th>BE776688</th><th>MY-19-B-0</th></td></td></td>	1.0 <td>592<td>12<th>BE776688</th><th>BE776688</th><th>MY-19-B-0</th></td></td>	592 <td>12<th>BE776688</th><th>BE776688</th><th>MY-19-B-0</th></td>	12 <th>BE776688</th> <th>BE776688</th> <th>MY-19-B-0</th>	BE776688	BE776688	MY-19-B-0
31 <td>48.2<td>1.0<td>367<td>13<th>Bj432209</th><th>Bj432209</th><th>Bj432209</th></td></td></td></td>	48.2 <td>1.0<td>367<td>13<th>Bj432209</th><th>Bj432209</th><th>Bj432209</th></td></td></td>	1.0 <td>367<td>13<th>Bj432209</th><th>Bj432209</th><th>Bj432209</th></td></td>	367 <td>13<th>Bj432209</th><th>Bj432209</th><th>Bj432209</th></td>	13 <th>Bj432209</th> <th>Bj432209</th> <th>Bj432209</th>	Bj432209	Bj432209	Bj432209
32 <td>48<td>1.0<td>559<td>9<th>AU093196</th><th>AU093196</th><th>AU093196</th></td></td></td></td>	48 <td>1.0<td>559<td>9<th>AU093196</th><th>AU093196</th><th>AU093196</th></td></td></td>	1.0 <td>559<td>9<th>AU093196</th><th>AU093196</th><th>AU093196</th></td></td>	559 <td>9<th>AU093196</th><th>AU093196</th><th>AU093196</th></td>	9 <th>AU093196</th> <th>AU093196</th> <th>AU093196</th>	AU093196	AU093196	AU093196
33 <td>48<td>1.0<td>680<td>12<th>BF620370</th><th>BF620370</th><th>HVSMC001</th></td></td></td></td>	48 <td>1.0<td>680<td>12<th>BF620370</th><th>BF620370</th><th>HVSMC001</th></td></td></td>	1.0 <td>680<td>12<th>BF620370</th><th>BF620370</th><th>HVSMC001</th></td></td>	680 <td>12<th>BF620370</th><th>BF620370</th><th>HVSMC001</th></td>	12 <th>BF620370</th> <th>BF620370</th> <th>HVSMC001</th>	BF620370	BF620370	HVSMC001
34 <td>48<td>1.0<td>818<td>10<th>BE040681</th><th>BE040681</th><th>OP04H09 O</th></td></td></td></td>	48 <td>1.0<td>818<td>10<th>BE040681</th><th>BE040681</th><th>OP04H09 O</th></td></td></td>	1.0 <td>818<td>10<th>BE040681</th><th>BE040681</th><th>OP04H09 O</th></td></td>	818 <td>10<th>BE040681</th><th>BE040681</th><th>OP04H09 O</th></td>	10 <th>BE040681</th> <th>BE040681</th> <th>OP04H09 O</th>	BE040681	BE040681	OP04H09 O
35 <td>47.8<td>1.0<td>674<td>9<th>AI820329</th><th>AI820329</th><th>605091F03</th></td></td></td></td>	47.8 <td>1.0<td>674<td>9<th>AI820329</th><th>AI820329</th><th>605091F03</th></td></td></td>	1.0 <td>674<td>9<th>AI820329</th><th>AI820329</th><th>605091F03</th></td></td>	674 <td>9<th>AI820329</th><th>AI820329</th><th>605091F03</th></td>	9 <th>AI820329</th> <th>AI820329</th> <th>605091F03</th>	AI820329	AI820329	605091F03
36 <td>47.6<td>1.0<td>621<td>17<th>BH688119</th><th>BH688119</th><th>hk22E04 Y</th></td></td></td></td>	47.6 <td>1.0<td>621<td>17<th>BH688119</th><th>BH688119</th><th>hk22E04 Y</th></td></td></td>	1.0 <td>621<td>17<th>BH688119</th><th>BH688119</th><th>hk22E04 Y</th></td></td>	621 <td>17<th>BH688119</th><th>BH688119</th><th>hk22E04 Y</th></td>	17 <th>BH688119</th> <th>BH688119</th> <th>hk22E04 Y</th>	BH688119	BH688119	hk22E04 Y
37 <td>47.6<td>1.0<td>647<td>13<th>BM654776</th><th>BM654776</th><th>170006873</th></td></td></td></td>	47.6 <td>1.0<td>647<td>13<th>BM654776</th><th>BM654776</th><th>170006873</th></td></td></td>	1.0 <td>647<td>13<th>BM654776</th><th>BM654776</th><th>170006873</th></td></td>	647 <td>13<th>BM654776</th><th>BM654776</th><th>170006873</th></td>	13 <th>BM654776</th> <th>BM654776</th> <th>170006873</th>	BM654776	BM654776	170006873
38 <td>47.2<td>1.0<td>471<td>10<th>BE356945</th><th>BE356945</th><th>DG1 145 F</th></td></td></td></td>	47.2 <td>1.0<td>471<td>10<th>BE356945</th><th>BE356945</th><th>DG1 145 F</th></td></td></td>	1.0 <td>471<td>10<th>BE356945</th><th>BE356945</th><th>DG1 145 F</th></td></td>	471 <td>10<th>BE356945</th><th>BE356945</th><th>DG1 145 F</th></td>	10 <th>BE356945</th> <th>BE356945</th> <th>DG1 145 F</th>	BE356945	BE356945	DG1 145 F
39 <td>47.2<td>1.0<td>478<td>10<th>BE583083</th><th>BE583083</th><th>3-10F-MF</th></td></td></td></td>	47.2 <td>1.0<td>478<td>10<th>BE583083</th><th>BE583083</th><th>3-10F-MF</th></td></td></td>	1.0 <td>478<td>10<th>BE583083</th><th>BE583083</th><th>3-10F-MF</th></td></td>	478 <td>10<th>BE583083</th><th>BE583083</th><th>3-10F-MF</th></td>	10 <th>BE583083</th> <th>BE583083</th> <th>3-10F-MF</th>	BE583083	BE583083	3-10F-MF
40 <td>47.2<td>1.0<td>464<td>14<th>BM834231</th><th>BM834231</th><th>K-EST0109</th></td></td></td></td>	47.2 <td>1.0<td>464<td>14<th>BM834231</th><th>BM834231</th><th>K-EST0109</th></td></td></td>	1.0 <td>464<td>14<th>BM834231</th><th>BM834231</th><th>K-EST0109</th></td></td>	464 <td>14<th>BM834231</th><th>BM834231</th><th>K-EST0109</th></td>	14 <th>BM834231</th> <th>BM834231</th> <th>K-EST0109</th>	BM834231	BM834231	K-EST0109
41 <td>47.2<td>1.0<td>576<td>14<th>BM846866</th><th>BM846866</th><th>K-EST0125</th></td></td></td></td>	47.2 <td>1.0<td>576<td>14<th>BM846866</th><th>BM846866</th><th>K-EST0125</th></td></td></td>	1.0 <td>576<td>14<th>BM846866</th><th>BM846866</th><th>K-EST0125</th></td></td>	576 <td>14<th>BM846866</th><th>BM846866</th><th>K-EST0125</th></td>	14 <th>BM846866</th> <th>BM846866</th> <th>K-EST0125</th>	BM846866	BM846866	K-EST0125
42 <td>47.2<td>1.0<td>592<td>10<th>BE584361</th><th>BE584361</th><th>10-ID-HA</th></td></td></td></td>	47.2 <td>1.0<td>592<td>10<th>BE584361</th><th>BE584361</th><th>10-ID-HA</th></td></td></td>	1.0 <td>592<td>10<th>BE584361</th><th>BE584361</th><th>10-ID-HA</th></td></td>	592 <td>10<th>BE584361</th><th>BE584361</th><th>10-ID-HA</th></td>	10 <th>BE584361</th> <th>BE584361</th> <th>10-ID-HA</th>	BE584361	BE584361	10-ID-HA
43 <td>47.2<td>1.0<td>597<td>13<th>BI818181</th><th>BI818181</th><th>603032640</th></td></td></td></td>	47.2 <td>1.0<td>597<td>13<th>BI818181</th><th>BI818181</th><th>603032640</th></td></td></td>	1.0 <td>597<td>13<th>BI818181</th><th>BI818181</th><th>603032640</th></td></td>	597 <td>13<th>BI818181</th><th>BI818181</th><th>603032640</th></td>	13 <th>BI818181</th> <th>BI818181</th> <th>603032640</th>	BI818181	BI818181	603032640
44 <td>47.2<td>1.0<td>609<td>10<th>BE583082</th><th>BE583082</th><th>3-9A-MY P</th></td></td></td></td>	47.2 <td>1.0<td>609<td>10<th>BE583082</th><th>BE583082</th><th>3-9A-MY P</th></td></td></td>	1.0 <td>609<td>10<th>BE583082</th><th>BE583082</th><th>3-9A-MY P</th></td></td>	609 <td>10<th>BE583082</th><th>BE583082</th><th>3-9A-MY P</th></td>	10 <th>BE583082</th> <th>BE583082</th> <th>3-9A-MY P</th>	BE583082	BE583082	3-9A-MY P

ALIGNMENTS

RESULT 1
LOCUS AZ289042
DEFINITION RPCI-23-59A6.TVB RPCI-23 Mus musculus genomic clone RPCI-23-59A6,
DNA sequence.
ACCESSION AZ289042
VERSION AZ289042.1 GI:9530917
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 489)
AUTHORS Zhao,S., Nieman,W., Felblyum,T., Malek,J., Shatsman,S., Akiret
,B., Levine,M., McGam,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Frazer,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
COMMENT Unpublished (1999)
Other_GSSs: RPCI-23-59A6.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from

ACCESSION	BH115587	
VERSION	BH115587.1	GI:14954954
KEYWORDS	GS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	

QY	AGAAACATCAATTTTGGTCAGACATCGGCATCTCAGAGTCAGTCCCGGATCCACAACTCTCG	218
Db	420 GTTATGCCAAAAACGATAGAAAGTGAGAGCGCTGAGCTCATTTAGTCTATGACAAACAGATGTA	479
QY	2788 GAGAACCTCCAGCAACCCCGCTGCTGTGGACCTTACTACATAGGCTTCAGAGCGGTGGCG	2847
Db	480 GCAGTTTCCAGACAGATATCCTCCAAATTAGCACTTCTATATCTCTGAGACTGGAGGTC	539
QY	2848 CACCAATGGCAGACATATACGAGAGCGCCGCGAGATGGTGTAATGGCTCAGGAAATTGGC	2907
Db	540 TACTAATGGGCAACAAACAC - CAGACACTGATGAGTGAGCAATGGCCATGATGATTGGC	598
QY	2308 ATTTCGATTCCATATGGCTGGGCGACAGAGTCATCACCACGAGACCCGCACTGGGCTT	296

Dn		599	AATGCGCATCCAAAGTGTAATGGTAGACTGACGTATTCACTTATCTTCGCCAGACTGGGTC	658
Oy		2368	TGCCCACTAACATAACCACCTTTACAAGAATCTCCA	3006
Dd		659	TGCCCACTCCAATCACATTTCACCAACTATATGCACA	697
RESULT 3				
LOCUS	AZ028362/c			
DEFINITION	RPCI-23-317G13.TV RPCI-23 Mus musculus genomic clone RPCI-23-317G13 DNA sequence.	495 bp DNA linear GSS 25-FEB-2000		
VERSION	AZ028362			
KEYWORDS	AZ028362.1 GI:7103746			
SOURCE	GSS. house mouse. Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 495)			
REFERENCE	Zhao,S., Niemman,W., Feldblum,T., Malek,J., Shataman,S., Akiret B., Levins,M., McGann,S., Tesgaye,G., Geer,K., Krol,M., de Jong,P. and Praser,C.M.			
AUTHORS	Mouse BAC End Sequences from Library RPCI-23 unpublished (1999) Other_GSSs: RPCI-23-317G13.TV			
TITLE				
JOURNAL				
COMMENT				

FEATURES	SOURCE	COMMENT
		Other GS8s: RPCT-23-317G13 .TV
		Contact: Shaying Zhao
		Department of Eukaryotic Genomics
		The Institute For Genomic Research
		9712 Medical Center Dr., Rockville, MD 20850, USA
		Tel.: 301 838 0200
		Fax: 301 838 0208
		Email: szhao@tigr.org
		Clones are derived from the mouse BAC library RPCT-23. For BAC
		library availability, please contact Pieter de Jong
		(pieter@dejong.med.buffalo.edu). Clones may be purchased from
		BAC/CPC Resources (http://bacopc.med.buffalo.edu/orderingframe.htm)
		or from Resea.ch Genetics (http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)
		Plate: 317 row: G column: 13
		Seq primer: SP6
		Class: BAC ends.
		Location/Qualifiers
		1..495

```

FEATURES
source
    location/Qualifiers
    1. .495
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPC1-23-317G13"
    /clone_id="RPC1-23"
    /sex="Female"
    /lab_host="DH10B"
    /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
    EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
    brain genomic DNA was isolated and partially digested
    with a combination of EcoRI and EcoRI MethyIase. Size
    selected DNA was cloned into the pBAC3.6 vector at the
    EcoRI sites. The ligation products were transformed into
    DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
124 a 109 c 92 g 170 t
ORIGIN

```

Query Match	1.5%	Score 71	DB 17	Length 495
Best Local Similarity	56.1%	Pred. No. 1.1e-06		
Matches 134	Conservative 0	Mismatches 105	Indels 0	Gaps 0

Qy	1031	GACCGGGGCATCACTCCGAGAGAGATGATTCAGAGACACAGCCTCGTACATCTCC	10950
Db	242	GAAAGGATATTCATTCAGAGAGAGATGCGTCTATAAAAACCAAGATCGAAACATCT	183
Qy	1091	TTCAACGCCGCTTCCCACTCGCGGCTCCCAATCAAGCGCGCTCTGGACATATGCCGCAAG	1150
Db	182	TACATATATCGCAGTCTTCCCAACCAAGATAAATCAAAAGGCGCTCTTGAACAAATGTACCAAG	123

QY 1151 ATCATGCGCGTGAACCAATCCGCCCGACATCACTGGTAAGGCCCCGCTCGCCCGCGAC 1210

Db 122 ATTGTGACTGTGACCAAGTAGTAGAAGATTAGCTGATTGGCAAAAGATCCACTGGCAGAT 63

QY 1211 ATTAAAAACCAACCGCATCTACCGCATCTCGAGCTGAACGGCTACGAACCTGCTACGGC 1269

Db 62 GTCCTCTCACTCAACATTACCAAGATTATTGGAGGAATGGCTATAGATCCAGTTTAAAC 4

RESULT 4	LOCUS	DEFINITION
BM440166	BM440166	636 bp mRNA linear EST 01-FEB-2002 pg1n.pk002.f22 Normalized Chicken Reproductive Tract cDNA Library

Intermediate filament binding protein, 500kD [Homo sapiens], mRNA sequence.

ACCESSION	BM440166
VERSION	BM440166.1
KEYWORDS	GI:18470941
SOURCE	EST.
	chicken. 11

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 636)

TITLE
JOURNAL
COMMENT

ESTs from Normalized Chicken Reproductive Tract cDNA library
University of Delaware and INRA, Tours-Poultry Unit Project
Unpublished (2002)
Contact: Larry A. Cogburn

Townsend Hall, Newark, DE 19717, USA
 • Tel: 302-831-1335
 Fax: 302-831-2832
 Email: cogburn@udel.edu, www.chickiest.udel.edu

FEATURES	Location/Qualifiers
source	1. .636

BASE COUNT	ORIGIN
103 a	216 c 220 g 89 t 8 others

Query Match	1.2%	Score 55.6	DB 13	Length 636
Best Local Similarity	46.1%	Pred. No. 0.0087		
Matches 224	Conservative 0	Mismatches 259	Indels 3	Gaps 1

[illegible]

Dy 1278 TCTTTCGCGTGAGCCCAAAAAGTTCCGGAGGCCGAACATCTGGCTGTTTGGGC 1337
| | | | |
Db 84 TGCGCAGTTCCGGACCCTGCACCGACGAGGCTGAGCTAACGACGAGCTGTGGCGC 143
| | | | |

1338 CCGGCACACACGGGCAAGACCAACATTCGGGAAGCATCGCCACCGCGTGCCTTCAAG 13
 144 GCTGCCGCAAGGACGAGGGGCGCTGCTCTGTCGCGGTGGGCGACCGCGCGGCC 203

QY 1398 GCTGCGTCACTGGACCAATGAGAACTTCCCTTCAATGATTTGCTGACAAAGTGGGA 1457
 Db 204 TCACCTTCCGCGGCTGGCGACGACGATACCTGCGAGACCTGCTCTGCGACGGTGA 263
 QY 1458 TCTGCTGGAGAGAGGAGCAAGTACGCGCAAGTCTGAGATTCGCCAAGCCATTCTCG 1517
 Db 264 TGAACAGAGGACACGCGCGCTGACGAGGGGCTGGCGTCCGTGAGAGGATCTCCA 323
 QY 1518 GCGGC--AGCAAGTGTGCGCTGACCAAAAGTCAAGTCTCGGCCAAGTCAACCCCA 1574
 Db 324 GGGACCTGAAAGATTCCTGAGAGGACCAAGCTGATGCGCGTCTCTGCACTTCCA 383
 QY 1575 CCCCCGTGATGTCACCTCCAAACCAACATGATGCGCGTATGACGGGAAACAGACCA 1634
 Db 384 CCAAGAGCGGCTCTCCGTCTACAGGCGCATGAAAGGAGCATATCCGCGCCGACCG 443
 QY 1635 CTTTCGAGACACAGACCGGTTGAGAGACCGGATGTTCAATTGAACTACCCGCGTC 1694
 Db 444 CTTTCGAGCTCTGAGGCGCGAGGCGGCGACGCGGTAAGTGTGACCCCATCAAGGGGC 503
 QY 1695 TGGAGC 1700
 Db 504 TGAAGC 509

RESULT 5
 BM371945 659 bp mRNA linear EST 23-JUL-2002
 LOCUS Ebma08_S0004_H16_R maternal, 28 DPA, no treatment, cv Optic, Ebma08
 DEFINITION Hordeum vulgare cDNA clone Ebma08_S0004_H16_5', mRNA sequence.
 ACCESSION BM371945
 VERSION BM371945.2 GI:21943561
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 1 (bases 1 to 659)
 Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)
 On Jan 10, 2002 this sequence version replaced gi:18115335.
 Contact: Waugh R, Marshall DF
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.
 Location/Qualifiers
 1..659
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="Ebma08_S0004_H16"
 /clone_lib="maternal, 28 DPA, no treatment, cv Optic,
 Ebma08"
 /tissue_type="maternal"
 /dev_stage="28 DPA"
 /lab_host="DH10B"
 /note="Vector: pSPORT1, Site 1: Sal I; Site 2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from maternal tissue dissected from developing
 grains (28 days post anthesis) in glasshouse grown barley
 plants. Developed as part of the barley transcriptome
 resources of BBSRC/SERAD funded cereal IGF (Investigating
 Gene Function) project."

BASE COUNT
 ORIGIN

167 a 252 c 176 g 64 t

Query Match 1.2%; Score 55.2; DB 13; Length 659;
 Best Local Similarity 47.6%; Pred. No. 0.011; Indels 3; Gaps 1;
 Matches 196; Conservative 0; Mismatches 213;

QY 1078 CTGTGATCTCTCTTCAACGCGCTTCCATCTGCGGTCCTCAGATCAAGGCGCTGTGA 1137
 Db 25 CGCGACCCCGGCTCTTCTCTCCCTCCCGCCCGCCCAACAAGCGCGCTTCCAA 84
 QY 1138 CAATGCCGGAATCATGCGCTGACCAATCCGCGCCGACTACTGG--TAGGCC 1194
 Db 85 GCCCAACACAGCTCAACGCGCCGCGCTCATCGGTCGCGGAAGTCCCTTCGCGC 144
 QY 1195 CGTCCGCGCGGAGCACTTAAACCAACGCTATCAACCGCATCTCGAGTGAACGCTA 1254
 Db 145 CACCAACACAGCAAGGCGCGACCGCTTCAACAACATGACCTCCCGACGCAAGCCCTA 204
 QY 1255 CGAAGCTGCTTACGCGGCTCGCTCTTCTCGGCTGGGCCAGAAAAAGTTGCGGAAGC 1314
 Db 205 CGCAACGCGCAAGCCCAAGCGCGGCTACGCGCGCAAGCCCAACAGCCCAATCAACG 264
 QY 1315 CAACACCATCTGCTGTTTGGCGCGGCGCACACGCGCAAGCAACATCGCGGAAGCAT 1374
 Db 265 CAACACCTTGGGCTCAGCAAGATGGCGCGCACGCGCGCGGCTTACAGCAACTTCAA 324
 QY 1375 CGGCCACGCGGCTCTTCAAGCGCTGCTCACTGGAACCAATGAGAACTTCCCTTCAA 1434
 Db 325 CGGCGGCAACAGAGGGGTCAAGAGCTACTTCAACAAGTCAATGCGAGCGCGCGCA 384
 QY 1435 TGATTGCTGCAAGAGTGTGATCTGTGGAGAGGCGCAAGATGACGCGC 1486
 Db 385 CAACAGCAACAGAGCAACTTCAACTTCAACGCGCTACGCGCGCAAGAGGCGC 436

RESULT 6
 BQ766288 692 bp mRNA linear EST 26-JUL-2002
 LOCUS EBro08_S0005_F13_R root, 3 week, drought-stressed, cv Optic, EBro08
 DEFINITION Hordeum vulgare cDNA clone EBro08_S0005_F13_5', mRNA sequence.
 ACCESSION BQ766288
 VERSION BQ766288.1 GI:21974762
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 1 (bases 1 to 692)
 Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)
 Contact: Waugh R, Marshall DF
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.
 Location/Qualifiers
 1..692
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBro08_S0005_F13"
 /clone_lib="root, 3 week, drought-stressed, cv Optic,
 EBro08"
 /tissue_type="root"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /note="Vector: pSPORT1, Site 1: Sal I; Site 2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from roots of 3 week old drought stressed barley

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

FEATURES
 source

1..692
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBro08_S0005_F13"
 /clone_lib="root, 3 week, drought-stressed, cv Optic,
 EBro08"
 /tissue_type="root"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /note="Vector: pSPORT1, Site 1: Sal I; Site 2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from roots of 3 week old drought stressed barley

ACCESSION BJ268130
VERSION BJ268130.1 GI:20088244
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
REFERENCE Ogiwara, Y. and Murai, K.
AUTHORS Expressed genes in Triticum aestivum
TITLE Unpublished (2002)
JOURNAL Contact: Tadasi Shin-i
COMMENT Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehin@genes.nig.ac.jp.
FEATURES
source
1. 652
Location/Qualifiers
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="who15f01"
/clone_1ib="Y. Ogiwara unpublished cDNA library, wh_oh"
/tissue_type="plastid at heading date"
/dev_stage="Feekes' scale 10.5"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid.
Site_1: EcoRI; Site_2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhunov
in V. Dvorak Lab). Total RNA was prepared from sheath
tissue, equal quantities of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo
excised to give Bluescript phagemids in the TV Close lab
at the University of California, Riverside (Akhunov, Chin
, Choi, Close, Fenton, Kitanian, Otto, Simons, Zhang).
Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."

BASE COUNT 166 a 209 c 189 g 85 t 3 others
ORIGIN

Query Match 1.1%; Score 54; DB 13; Length 652;
Best Local Similarity 49.3%; Pred. No. 0.022;
Matches 141; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 1384 CGTCCCTTCTACGCGCTGCTCAACTGGAACCAATGAACTTCTTCATGATTGCT 1443
DB 237 CATACACTGCGACGCGCTGCAAGCAAGGTCAAGAGCTGCTTCAAGAGATCGAAGCGT 296
QY 1444 CGACAAGTGTGATCTGTGTGGAGAGAGGCGCAAGTACCGCAAGTCTGTGAGTCCG 1503
DB 297 CTACTCGGTGCGATCGATGTCACACCAAGGTCAAGTACCGCGACGCTGACTC 356
QY 1504 CAAGGCACTTCTGCGCGCGACGACAGGTGCGGTGACCAAAAGTGAAGTGTGCGCCCA 1563
DB 357 TGAAGCGCTATCTCGGAAGCTCACCAAGAGGCGGCAAGACGCAAGCTGTGTGACCA 416
QY 1564 GATGACCCCAACCCCGTGTGTCACCTTCAACCAACCAAGTGTGCGCGTGTGATGACG 1623
DB 417 GAAAGGCGGCAACGCGCGCAACCAACCAAGGCGGCGGCGCGCGCGGCGG 476
QY 1624 GAAACAGCACCACTTTCAGACGACGACCGCTTTCAGACGACGAGT 1669
DB 477 CAACAAACCAACGACGACGACGACGACGACGACGACGACGACGACGACG 522

RESULT 9
B0660503 B0660503 607 bp mRNA linear EST 15-JUL-2002
LOCUS
DEFINITION H102005u HI Hordeum vulgare cDNA clone H102005 3-PRIME, mRNA

sequence.
ACCESSION B0660503
VERSION B0660503.1 GI:21801636
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
REFERENCE Michael, W., Wesche, W., Pleisner, K.-P. and Graner, A.
AUTHORS EST sequencing and analysis in barley
TITLE Unpublished (2000)
JOURNAL Contact: Stein Nils
COMMENT Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert length: 607 Std Error: 0.00
Plate: 2 row: 0 column: 5
Seq primer: M13uni.
FEATURES
source
1. 607
Location/Qualifiers
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="GABI:136895"
/db_xref="taxon:4513"
/clone="H102005"
/clone_1ib="H1"
/tissue_type="female inflorescences"
/dev_stage="female inflorescences (approx. 3 mm in size)"
/lab_host="XL10-Gold"
/note="Vector: Bluescript SK-; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable."

BASE COUNT 132 a 251 c 162 g 62 t
ORIGIN

Query Match 1.1%; Score 53.2; DB 14; Length 607;
Best Local Similarity 47.9%; Pred. No. 0.034;
Matches 185; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

QY 1078 CTGTACATCTCTTCAACCGCGCTTCAACTGCGGTGCCAGATCAAGCGCGCTTGA 1137
DB 205 CGCGACCGCGGCTCTTCTCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 264
QY 1138 CAATGCCGGAATCATGCGCTGACCAATCCGCGCGCGCGCGCGCGCGCGCGCGCG 1194
DB 265 GCCCAACAAAGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 324
QY 1195 CGCTCGCGCGCGGCACTTAAACCAACCGCATCTACCGCATCTGAGTGAACGCGCTA 1254
DB 325 CACCAACACGACGAGCG 384
QY 1255 CGAAGCTGCTACGCGCGCTCGCTCTTCTTCTGCGGTGCGCGCGCGCGCGCGCGCG 1314
DB 385 CGCAAGCGCAACG 444
QY 1315 CAACACCTTGTGCTGTTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1374
DB 445 CAACACCTTGTGCGCTGACGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504
QY 1375 CGCCACGCGCGCGCTTCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1434
DB 505 CGGCGCGCAAGAGGGGTCAAGAGCTTCTTCAACCAAGTCCATGCGCGCGCGCGCGCG 564

QY 1435 TGATTCGTCGACAGATGCTGATCT 1460
 Db 565 CAACAGCAACAGCAGCAACTTCAACT 590

RESULT 10
 AL818251 528 bp mRNA linear EST 15-JUL-2002
 LOCUS AL818251 j:223 Triticum aestivum cDNA clone C05_j223_plate_14, mRNA
 DEFINITION sequence.
 ACCESSION AL818251 GI:21829851
 VERSION AL818251
 KEYWORDS EST.
 SOURCE Triticum aestivum
 ORGANISM bread wheat.
 REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticaceae; Triticum.
 1 (bases 1 to 528)
 Wilson, I., Beswick, R., Shepherd, S., Barker, G., Parker, J., Owen, P., Edwards, K., Coghill, J., Holdsworth, M., Lenton, J., Shewry, P. and A BBSRC-funded wheat EST resource for the academic community
 JOURNAL Unpublished (2002)
 COMMENT Contact: Barker G
 Institute of Arable Crop Research
 Long Ashton, Bristol BS41 9AF United Kingdom.
 FEATURES
 source
 1. 528
 /organism="Triticum aestivum"
 /cultivar="mercia"
 /db_xref="taxon:4565"
 /clone="C05_j223_plate_14"
 /clone_1lb="j:223"
 /tissue_type="embryo"
 /dev_stage="21 days post anthesis"
 /note="6 and 10 hour drought stress by placing plants on moist paper (75% rel. humidity) in light"
 BASE COUNT 138 a 169 c 157 g 64 t
 ORIGIN

Query Match 1.1%; Score 52.4; DB 9; Length 528;
 Best Local Similarity 49.0%; Pred. No. 0.052;
 Matches 140; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1384 CGTGGCCCTTACGCGTGCCTCACTGACCAATGAGAACTTCCCTTCATGATGGCGT 1443
 Db 195 CATACACTGCGAGCGCTGCAAGCAAGGTCAAGAACTCTCTCAAGATCTCAAGCGCT 254
 QY 1444 CGAACAAGTGTGATCTGTGGAGAGGCGCAAGATGACCGCCAAAGTCTGTGAGTCCG 1503
 Db 255 CTACGTGGTGGCATGATGTGGCAACCAAGGTCAAGGTGACGGGCGAGGTGAGCTC 314
 QY 1504 CAAGGCACTTCTGCGCGCGCAAGGTGCGCGTGTGACCAAAAGTCAAGTCTCCGCCA 1563
 Db 315 TGAAGCGCTGATCCGGAAGCTCAACAGAGGGGCAACAGAGCTGTGTGCAACA 374
 QY 1564 GATCGAAGCCGACCCGCGTGTGATCTGATCTCAACCAACATGATGCGCGTGTGACCG 1623
 Db 375 GAAGGGCGGCAAGCGCGCAACAACACTACAGGCTCAAGGCGCGCGCGCGCGG 434
 QY 1624 GAACAGACCACTTGTGAGCAGCAGCAGCGCTTGTGAGACCGGATG 1669
 Db 435 CAACCAACCAACAGAGAGCGCGAGCAGCAGCAGCAGCAGCAGAG 480

RESULT 11
 BM816637 688 bp mRNA linear EST 05-MAR-2002
 LOCUS BM816637
 DEFINITION HB02806.T3.abi HB Hordeum vulgare cDNA clone HB02806.T3.abi similar to Arginine decarboxylase, Arginine 2-monooxygenase, Glycine dehydrogenase (decarboxylating), mRNA sequence.
 ACCESSION BM816637
 VERSION BM816637.1 GI:19152651
 KEYWORDS EST.

SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticaceae; Hordeum.
 1 (bases 1 to 688)
 Ozturk, N.Z., Michalowski, C.B., Brazille, S., Borchert, C., Palacio, C., Normand, C., Murphy, C., Kelley, R., Sant, S.A., McLaughlin, H., Fredricksen, M.A. and Bohner, H.J.
 Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley
 JOURNAL Unpublished (2002)
 COMMENT Contact: Mark A. Fredricksen
 Plant Biology
 University of Illinois
 1201 W Gregory Dr, Urbana, IL 61801, USA
 Tel: 2172655473
 Email: bohner@life.uiuc.edu.
 FEATURES
 source
 1. 688
 /organism="Hordeum vulgare"
 /strain="cv Tokak"
 /db_xref="taxon:4513"
 /clone="HB02806.T3.abi"
 /clone_1lb="HB"
 /tissue_type="leaf"
 /dev_stage="3 week old"
 /note="6 and 10 hour drought stress by placing plants on moist paper (75% rel. humidity) in light"
 BASE COUNT 181 a 258 c 157 g 89 t 3 others
 ORIGIN

Query Match 1.1%; Score 52.4; DB 14; Length 688;
 Best Local Similarity 48.1%; Pred. No. 0.057;
 Matches 180; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

QY 1078 CTGCTACATCTCTTCAACGCGCTTCAACTGCGCGTCCAGATCAAGCGCGCTTGA 1137
 Db 153 CGCGAAGCCCGGCTCTTCTTCCCTCCCGCGCGCGCGCGCAACAAGCGCTTCCAA 212
 QY 1138 CAATGCGGCAAGATCATGCGCTGACCAATCGCGCGCGCTACTGCG--TAGGCCC 1194
 Db 213 GCCCAACAACAGCTCAACGCGCGCGCGCTCAATCGGCTCGGCAAGCTCGCTTGGCGC 272
 QY 1195 CGCTCGCGCGCGGAGATTAACAACGCGCATCTACCGCATCTCTGAGTGAACGCGTA 1254
 Db 273 CACCAACACCAAGAGCGCGAGCGCGCTCAACAACGCTACCGCTCCCGAGCGCAAGCCCTA 332
 QY 1255 CGAAGCTGCTAGCGCGGCTCGCTTCTTCTCGGCTGGCGCCAGAAAGTTGCGGAAGCG 1314
 Db 333 CGCAACGCGCAAGCGCGCTGAGCGCGCGCGCGCGCAAGCGCGCAACATCAACG 392
 QY 1315 CAACACATCTGCTGTTTGGCGCGCAACAGCGCGCAAGCAACATGCGCGAAGCCAT 1374
 Db 393 CAACACTTTCGGCTCAACAAGATGCGCGCGCGCGCGCGGCGGCTACAGCAACTTCAA 452
 QY 1375 CGCCACGCGCGCTTCTTGAAGCTGCTCAACTGAGCAACATGAGAACTTCCCTCAA 1434
 Db 453 CGCGGCAACAGAGGGGCTCAAGACTTCAACAAGTCAATCGGACGCGCGCGCA 512
 QY 1435 TGATTCGTCGACA 1448
 Db 513 CAACAGCAACAGCA 526

RESULT 12
 AY110092/c 1065 bp mRNA linear HTC 25-MAY-2002
 LOCUS AY110092
 DEFINITION Zea mays Cl49602_1 mRNA sequence.
 ACCESSION AY110092
 VERSION AY110092.1 GI:21214182
 KEYWORDS HTC.
 SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

AUTHORS 1 (bases 1 to 1065)

TITLE Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.

JOURNAL Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes

REFERENCE Unpublished (2002)

AUTHORS 2 (bases 1 to 1065)

JOURNAL Direct Submission

TITLE Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

FEATURES

source

1. 1065

/organism="Zea mays"

/db_xref="taxon:4577"

/clone="CLA9602_1"

/clone_lib="Maize Mapping Project/Dupont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"

BASE COUNT 178 a 300 c 260 g 221 t 106 others

ORIGIN

Query Match 1.1%; Score 52.4; DB 11; Length 1065;

Best Local Similarity 43.7%; Pred. No. 0.065;

Matches 204; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

Qy 1308 GGAAGCCGACACCATCTGCTGTTGGCGCGCACACGGGCAAGCAACATCGCG 1367

Db 916 GCAAGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCCCGCTCGATGACCCGG 857

Qy 1368 AAGCCATGCCCGCGCGCTTCTTACGCGCTGCTGCTGCACTGACCAATGAACTTTC 1427

Db 856 AGGAGCGCGATGCTGCGCTCCCGACGCGGAGCTGACAGTTCACCGCTTCTTCC 797

Qy 1428 CTTCAATGATTTGGCTGACAAAT--GGTATCTGGTGGAGAGGCGCAATGACG 1484

Db 796 TCGAAGAGAGAGACTACGCTATCGGCTGAAGATTCAGAGAGAGGTGTGACG 737

Qy 1485 CCAAGTCTGTGAGTCCGCCAAGGCAATCTCGCGCGAGCAAGTGGCGCTGACCAA 1544

Db 736 CCGGAGAGTGGGGTGGCGGAGAGCTGCGGGGTGGCAAGATCTGCACTTCC 677

Qy 1545 AGTCGAAGTCTCGCGCCGATCGACCCACCCCGTATCTCACTTCAACCAACA 1604

Db 676 ACGCGAGATGCTCTTGTGAGAACTACACGCGCTCAACACCGGCTGTGTAAGA 617

Qy 1605 TGTGCGCGGTATGAGCGGAACAGACCACTTCAAGACAGACCGCTTGAAGACC 1664

Db 616 TCTCAAGAAATGACAAAGAGACCGGCGGTGATCGCTTCAATCGGAACG 557

Qy 1665 GGATGTTCAATTTGAATCAACCGCGCTTGAAGCATGATTTGGCAAGGTGACAAAGC 1724

Db 556 TGATGCGAGAGCCGTTCTGCGCCACGACGCTGTCAACAGTCTGTAAGGGGTGCAAG 497

Qy 1725 AGGAATCAAAAGATTCTTCGCTGGGCGCAGATCACTGACGAG 1771

Db 496 AGATGCTGAGACCACTCTCTGCGCCGGAACATCAAGCGGCGGTG 450

RESULT 13

LOCUS B1952167 877 bp mRNA linear EST 19-OCT-2001

DEFINITION HVSMEM0004N17E Hordeum vulgare green seedling EST library.

ACCESSION HVCDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone

VERSION B1952167

KEYWORDS B1952167.1 GI:16295739

SOURCE EST.

ORGANISM Hordeum vulgare.

REFERENCE Hordeum vulgare

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae / Triticeae; Hordeum.

TITLE 1 (bases 1 to 877)

JOURNAL Wing, R., Close, T.J., Kleinbofs, A., Wise, R., Chin, A., Begum, D., Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and Main, D.

COMMENT Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library

Unpublished (2001)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 467

Seq primer: AATTACCTCACTAAGG

High quality sequence stop: 843.

Location/Qualifiers

1. 877

/organism="Hordeum vulgare"

/cultivar="Morex"

/db_xref="taxon:4513"

/clone="HVSMEM0004N17E"

/clone_lib="Hordeum vulgare green seedling EST library

HVCDNA0014 (Blumeria infected)"

/tissue type="green seedling leaf"

/lab host="TTC121"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Morex (mla) plants were greenhouse grown in the R wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinbofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 192 a 319 c 259 g 105 t 2 others

ORIGIN

Query Match 1.1%; Score 52; DB 13; Length 877;

Best Local Similarity 47.1%; Pred. No. 0.077;

Matches 194; Conservative 0; Mismatches 215; Indels 3; Gaps 1;

Qy 1078 CTGTACATCTCTTCAACGCGCTTCAACTCGCGCTCCAGATCAAGCGCGCTGTGA 1137

Db 210 CGCCGACCCGGGCTTTCTCCCTCCCGCCGCGCCGACAAACAGCGCTTCCAA 269
 Qy 1138 CAATGCGGCAAGATCATGCGGTGACCAATCCGGCCCGGACTACTGG---TAGGCC 1194
 Db 270 GCCCAACAAACAGCTTCAACGCGCCCTCATGCGGTGCGGCAAGTCCCTTCGCGC 329
 Qy 1195 CGCTCGCGCCGCGGACATTAACCAACCGCATCTACCGCATCTGAGGTGAAACGCTA 1254
 Db 330 CACCACCAACGACGAGCGCCACCGCTACACAGCTACGCTCCCGACCGCAAGCTTAA 389
 Qy 1255 CGAACCTGCTACGCGCGCTCCGCTTTCTTCGCGTGGGCCCGAAGAAAGTTGCGAAGC 1314
 Db 390 CGCAACGCGCAACGCGCAACCGCGCTACGCGCCACCAACGCAACATCAACG 449
 Qy 1315 CAACACCATCTGCTGTTTGGCGCGGACCAACGCGGCAACCAATCGGAGAGCAT 1374
 Db 450 CAACACCTTGGGCTGCAAGATGCGCGCGCGCGCGCGCTACAGCACTTCA 509
 Qy 1375 CGCCACGCGCGTGCCTTTCTACGCGTGGTCACTGACCAATGAGAACTTCCCTTCA 1434
 Db 510 CGCGCGCAACGAGGGGCTCAAGAGCTTCAACAAATCCATCGCGCGCGCGCAAC 569
 Qy 1435 TGAATGCGTCAACAGATGCTGATCTGATGCGAGAGCGCAAGATGACGCGC 1486
 Db 570 CAGACGCAACGACGCAACTTCAACGCGCTACGCGCGCGCAAGAGGCG 621

RESULT 14
 BE452865 570 bp mRNA linear EST 26-JUL-2000
 LOCUS 894066D03.Y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BE452865
 VERSION BE452865.1 GI:9459700
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii;
 Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 570)
 Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
 McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 Unpublished (2000)
 CONTACT: Elizabeth H. Harris
 DCMB Box 91000
 Durham, NC 27708-1000, USA
 Tel: 919 613 8164
 Fax: 919 613 8177
 Email: chlamy@duke.edu.

JOURNAL COMMENT
 Location/Qualifiers
 1..570
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

FEATURES
 source

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described

BASE COUNT 130 a 182 c 167 g 91 t in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
 Query Match 1.1%; Score 51.6; DB 10; Length 570;
 Best Local Similarity 46.0%; Pred. No. 0.085;
 Matches 174; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

Qy 1332 TTGGGCGGCGACACGAGGGAAGACCAACATCGCGAAGCATCCGCGTCCCT 1391
 Db 181 TTGGCAAGAGGGCGAGATCAACACCGTCTTGAGCTACCTGCGCAACTTCTGCTGC 240
 Qy 1392 TCTACGCTGCTGCACTGGAACATGAGAACTTCTCCCTTCAATGATGCGTGCACAGA 1451
 Db 241 CCAACGGCATGCGCAAGATTGCCAGGAGGACCTCTCAACCAATCCGCGCAAGAG 300
 Qy 1452 TGGTGAATCTGTTGAGAGAGGAGGCAAGATGACGCGCAAGTGTGAGTCCGCAAGGCCA 1511
 Db 301 AGGACGAGATCCGCAAGAAAGCTGAGAGAGAGGCCAGGCGGAGGCTTTGCCAAGGCC 360
 Qy 1512 TTCTCGCGCGACAGCATGTCGCGTGGACCAAAAGTGCAGATGCTCCGCAATGCAGC 1571
 Db 361 TGACCAACATCGGCAAGTTCTGCTGAAGAGAACCGCGGAGAAAGACAGATCTACG 420
 Qy 1572 CCACCCCGGTGATCGTCACTCCACCAACATGTGCGCGGTGATTGACGGGACAGCA 1631
 Db 421 GCTCCGTGACGGTGCAGAGAGATTGCCAGGCCATCTTACAGAGACCGCGCGCAAGTGA 480
 Qy 1632 CCACCTTGCAGACACGACGACCGCTTCAGAGCCGATGTTCAATTGAACTACCGCGCC 1691
 Db 481 GCACAGTGCAGATCACTGTGTCGCGAGATCAAGTCGGTGGCACCTAGATGACCATTC 540
 Qy 1692 GTCTGAGCATGACTTGG 1709
 Db 541 GCCTGACACCCCGAGGTTG 558

RESULT 15
 B1873856 612 bp mRNA linear EST 11-OCT-2001
 LOCUS 963111F11.Y1 C. reinhardtii CC-1690, stress condition I, normalized
 DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION B1873856
 VERSION B1873856.1 GI:16071860
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii;
 Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 612)
 Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,
 P., McDermott,J.P., Shryger,J., Silflow,C. and Stern,D.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 963
 Unpublished (2001)
 CONTACT: Charles Hauser
 DCMB Box 91000
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.

JOURNAL COMMENT
 Location/Qualifiers
 1..612
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, stress condition I,
 normalized, Lambda Zap II"

FEATURES
 source

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 10.5714 Seconds

(without alignments)
2095.115 Million cell updates/sec

Title: US-09-807-802A-17

Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRVLTPL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	74.9	504	1 COA3 AAV2	P03135 adeno-assoc
2	517	17.8	673	1 COAT PAYBO	P07237 bovine parv
3	487.5	16.8	781	1 COAT PAYHB	P07239 human parvo
4	235.5	8.1	749	1 COA2 PAYVN	P22964 porcine par
5	232.5	8.0	729	1 COA1 PAYP9	P33484 porcine par
6	232.5	8.0	729	1 COA1 PAYVN	P18546 porcine par
7	230.5	7.9	729	1 COA1 PAYV3	P52501 porcine par
8	229	7.9	587	1 COAT PAYVH	P36310 parvovirus
9	226	7.8	722	1 COAT PAYVH	P03136 hamster par
10	197	6.8	716	1 COAT MIMIV	P03137 murine minu
11	196.5	6.8	716	1 COAT MIMIV	P07302 murine minu
12	194	6.7	722	1 COAT MEVA	P27437 mink enteri
13	194	6.7	727	1 COAT FPV	P04864 feline panl
14	190	6.5	584	1 COAT PAYC2	P30129 canine parv
15	189	6.5	727	1 COAT FPV19	P24840 feline panl
16	188	6.5	727	1 COAT PAYCB	P11213 canine parv
17	188	6.5	737	1 COAT PAYCB	P17455 canine parv
18	184	6.3	748	1 COAT PAYCN	P21930 canine parv
19	177.5	6.1	647	1 COAT ADVG	P24029 aleutian m
20	176	6.1	722	1 COAT PAYC7	P04863 canine parv
21	122.5	4.2	648	1 WHN HUMAN	O15333 homo sapien
22	121	4.2	880	1 SYV BACST	P11931 bacillus st
23	109	3.8	1186	1 CEAA BACST	O45710 bacillus th
24	108	3.7	655	1 CDGT KLEPN	P08704 klebsiella
25	107.5	3.7	648	1 WHN MOUSE	O61575 mus musculu
26	106.5	3.7	1070	1 Y355 HUMAN	O15063 homo sapien
27	106	3.6	1271	1 Y338 MYCBE	P47580 mycoplasma
28	105	3.6	1396	1 ITA2 DROME	P12080 drosophila
29	104.5	3.6	559	1 HNEB PIG	O03365 sus scrofa
30	104.5	3.6	717	1 CLAI ARATH	O38864 arabidopsis
31	104	3.6	395	1 MAFI YEAST	P41910 saccharomyc
32	103.5	3.6	1196	1 AMYB PAEPO	P21543 paenibacill
33	102	3.5	1122	1 ADPI MYCGA	Q49379 mycoplasma

34	102	3.5	2493	1 CYAA USTMA	P49606 ustilago ma
35	102	3.5	3712	1 IMA DROME	O00174 drosophila
36	100.5	3.5	531	1 HNEB PIG	O29548 sus scrofa
37	100	3.4	729	1 CATR MYGSM	O59557 mycobacteri
38	99	3.4	879	1 GUNI CLOTM	O02934 clostridium
39	98.5	3.4	557	1 HNEB RAT	P23899 rattus norv
40	97.5	3.4	560	1 GAT2 YEAST	P40209 saccharomyc
41	97.5	3.4	676	1 VGP EBOIC	O66810 ebola virus
42	97.5	3.4	685	1 RNFE HUMAN	O9Y252 homo sapien
43	97.5	3.4	1113	1 N116 YEAST	O02630 saccharomyc
44	97.5	3.4	2233	1 COAC YEAST	O00955 saccharomyc
45	97	3.3	800	1 GUN BACSI	P06564 bacillus sp

ALIGNMENTS

RESULT 1	ID	COA3 AAV2	STANDARD;	PRT;	504 AA.
AC	P03135;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-NOV-1991 (Rel. 20, Last annotation update)				
DE	Probable coat protein 3.				
OS	Adeno-associated virus 2 (AAV2).				
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=10804;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=63164299; PubMed=6300419;				
RA	Srivastava A., Luby E.W., Berns K.I.;				
RT	"Nucleotide sequence and organization of the adeno-associated virus 2				
RL	J. Virol. 45:555-564 (1983).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: J01901; AAA42376.1; -.				
DR	PIR: A03698; VCPV3A.				
DR	InterPro: IPR001403; Parvo_coat.				
DR	Pfam: PF00740; Parvo_coat; 1.				
KM	Coat protein.				
SQ	SEQUENCE 504 AA; 56366 MW; 758999B017052B6B2 CRC64;				
Query Match	74.9%; Score 2177.5; DB 1; Length 504;				
Best Local Similarity	82.5%; Pred. No. 5.7e-149;				
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;					
QY	1 MASGGAPMADNNEGADGVGNSGNHCHDSTWLGDRVITTSPTMTALPTNNHLYQGIS 60				
DB	1 MASGGAPMADNNEGADGVGNSGNHCHDSTWLGDRVITTSPTMTALPTNNHLYQGIS 60				
QY	61 ASTGASNDNHYFYSTPMGTFDPRNPHCHSPRDMORLINNMGFPKRLNPLFIQVK 120				
DB	61 QS-GASNDNHYFYSTPMGTFDPRNPHCHSPRDMORLINNMGFPKRLNPLFIQVK 119				
QY	121 EYTTNDGVTIANNLTSTVQVPSDSYQI.PYVLGSAHQGCLPPFPADVMTPOYGYLTIN 180				
DB	120 EYVQNGTITIANNLSTVQVFTDSYQI.PYVLGSAHQGCLPPFPADVMTPOYGYLTIN 179				
QY	181 NSGQAVGRSPFLCYEFPSPOMLRTGNNTFSTFEVPPHSSVYASQSLDRLNPLIDQY 240				
DB	180 NSGQAVGRSPFLCYEFPSPOMLRTGNNTFSTFEVPPHSSVYASQSLDRLNPLIDQY 239				
QY	241 LYYLTNFTQNSGSAONKDLLFSRGSFAGMSVOPKXNLPPCYRQGRVSKTKTDNNNSNFT 300				

Db 240 LYLSTNTSPGTTQSLQPSQAGASIDRDSRWMLPGPCYRQORVSKTSADNNNSEYS 299
 Qy 301 WTGSKYNLNGRESIIINPGTAMASHKODEDFPFMSGIMTIGESASASNTALNNMTD 360
 Db 300 WTGATKTHLNGRDLNVP--AMASHKODEEKFPGSGVLIGKQSEETVNIKWTMD 357
 Qy 361 EEEIKATNPVATFEFGVAVNFQSSSTDPAATGVHAMGALPGAWODRDVYLQGPIMAKI 420
 Db 358 EEEIGTNPVATEQVGSVSTLQKGNQATADVNTQGVLPQWQODRDVYLQGPIMAKI 417
 Qy 421 PHTDGHFHPSPLMGFGFKLPKPPOLLIKNPVNPANPPAEFSATKPAFTTQYSTG 475
 Db 418 PHTDGHFHPSPLMGFGFKLPKPPOLLIKNPVNPANPPAEFSATKPAFTTQYSTG 472

RESULT 2

COAT_PAVBO STANDARD; PRT; 673 AA.
 ID COAT_PAVBO STANDARD; PRT; 673 AA.
 AC P07297; O84374;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP2 [Contains: Coat protein VP3].
 OS Bovine parvovirus (BPV).
 CC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=10784;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87061184; PubMed=3783814;
 RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,
 RA Bates R.C.;
 RT "Complete nucleotide sequence and genome organization of bovine parvovirus."
 RL J. Virol. 60:1085-1097(1986).

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M14363; AAB59848.1; -
 DR EMBL; M14363; AAB59849.1; -
 DR PIR; A26104; VCPV85.
 DR HSP; P30129; ADPV.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 673
 FT CARBOHYD 138 673 COAT PROTEIN VP2.
 FT CARBOHYD 343 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 163 187 GLY-RICH.
 SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 17.8%; Score 517; DB 1; Length 673;
 Best Local Similarity 25.1%; Pred. No. 1.2e-29;
 Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;

Qy 5 GGAMADNNEGADGVNAGNWCDSWLGKRVTTSTRTVALPTYNHLYKQISSASTG 64
 Db 163 GSGGVGGGGRGSGVSTGWTGTTIFSENIIVTKTKRPFICDIKNGHLYKS-EVNTG 221
 Qy 65 ASDNHHYGVSTPMGPNRFGHCFSPRDMQRLINNMWGRPKRLNFKLNIQKWT 124
 Db 222 DTAHRQY-ATTPMSYFENFQYSSHFSPNDQMLHDERFRPKAMIVRYNLDIQKWT 280
 Qy 125 NDGVTTIANN-LTSTVQVFSDEYQLPYVLGSAHQGLPPADVFMIPQYGLTL- 179

Db 281 DGAMGTVYNNDLTLAGNHIIFCDGDHRYPYVQHPPMDQCMPELPMSISWELPGYATIPAPISV 340
 Qy 180 --NNGSAVRS-----SFLCEYFSPQMLRTGNFTFTFEVPHSSAASQSLDR 232
 Db 341 VDNNTNTVTEHLLKGVPLTMSDHEVLNNG-----RI 375
 Qy 233 MNPLIDQVLYLNRTON-----QSGSAONKDLFSESGP-----AGMSVQPGMWL 277
 Db 376 YRIYQLMRLMRKQHIIQHASDDVOSTGOKNLLIQTKQPNKORFQNALRTSNM 435
 Qy 278 PGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDFPFMSG 337
 Db 436 SGP-----GIARQTHNATLQTSAGALVTMTWTCAD-----VSG 469
 Qy 338 VM-----IFGKE-----SAGSNTLNDVNMITDEEIKATNPAT 372
 Db 470 VRAVRGVSTDPYGGQGPESDILLRLYSASAEGQNPILN-----AAR 515
 Qy 373 ERFGTVAVNFQSSSTDPAATGVHAMGALPGWQODRDVYLQGPIMAKIPHTDGHFHPSP 432
 Db 516 HTFRKARTLITGSCAGDGDYKEMWMLPNQMDSPASIKYNIWYKVPVNRKTLIDQ 575
 Qy 433 MGFGLKNPPOLLIK--NTPVNPANPPAEFSATKPAFTTQYSTQGVSEIEMLOKENS 490
 Db 576 DGSIPMSHPGTIFIKLARIPEVNGD-----SFLNIYTVQGVSCVWVEVERGT 626
 Qy 491 KRNPPEVQYTSNNAKSAND-FTVDNNGLY 519
 Db 627 KNMRPEYMS--ATNMSVDATYTNAGVY 653

RESULT 3

COAT_PAVHB STANDARD; PRT; 781 AA.
 ID COAT_PAVHB STANDARD; PRT; 781 AA.
 AC P07297;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DE Probable coat protein VPI.
 OS Human parvovirus B19.
 CC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
 NCBI_TaxID=10798;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate AU;
 RX MEDLINE=86200451; PubMed=3701931;
 RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;
 RT "Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis."
 RL J. Virol. 58:921-936(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M13178; AAA66867.1; -
 DR PIR; A24299; VCPV19.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576807 CRC64;

Query Match 16.8%; Score 487.5; DB 1; Length 781;
 Best Local Similarity 27.4%; Pred. No. 2e-27;

Matches 143; Conservative 84; Mismatches 239; Indels 55; Gaps 16;

QY 10 ADNNBAGADVGNAS--GNHCDSTWLGDRITTTSTWMLPTNNHLYKQISSASAGSN 67
 DB 234 AEASTAGAGGSGSVKSMSEGAFTSANSVTCTFSHQFLPYDEHHYKFSAPASCHN 293
 QY 68 D-----NHYFGSTPMGYFDFNRFHCHSPBDWQRLINNNGFRPKRLNFKLENT 117
 DB 294 ASGEKAKVCITSPIMKYSTFWRLDRLNALNLFSPLEFQILINYSIAPDALTYITSEI 353
 QY 118 QYKEVT--TNDVTTIANNLTSTVOVFSDEYQLPVYLSAHOGLCPPPADVFMIPQY 175
 DB 354 AVMDVDTKGGGV--QVTDSTTGRLCMVDHEKYPVYLGCGQDPTLAPBLPIWYFPPQYA 412
 QY 176 YLTLLN--GSGAVG-----RSSPYCLEFPSPQMLRGNNTFTSYTPEVEFHSYAH 225
 DB 413 YLTGVNTVQGISDGSKCLASESAFVLEHSSFQLGTGTGTSMSKFPPEVPEMLEGC 472
 QY 226 SOSLDRLANPLIDQVLYILNRTQNOGSAQNKDLFSRGSAPAGMSYQPKMLPGPCYRQ 285
 DB 473 SCHFYEMNPL--YGSRLGVPDTLGGDPKFRSL-----THEDHALQPOKFMGPLVNSV 524
 QY 286 RVSKTYTNNNSNFTWTGASKNLNGRESIINPG-TAMASHKODEKFFPMGSMVIFGKE 344
 DB 525 STEKGDSNTGAGKALTGLSTGTSQWTRISLRPGPVSQPYHMDTKYVGINAISHGQT 584
 QY 345 SAGASNTALDNV-----MITDEBEIKATNPVATERFGTVAVNPOSSSTDPATGVHAM 397
 DB 585 TVG--NAEDKEYQGVGRPEPNEKOLQLOGLNMTY-----FPKKGYYQYTDQIE-R 634
 QY 398 GALPGWODRDVYLOGPIWAKIPHTDGHFHS-PLMGFGGLKNPPQILIKTPVPAV 456
 DB 635 PLMGVSMVRRAALHYSQMSKIPLNDSPKTKQPALGCMGLMQPPQITLK--ILPQSG 652
 QY 457 PAEFSATKFSFTYQSTGVSEIEMEL--OKENSKRWPE 496
 DB 693 PIGGKSMGITTLYQVAVGIMVTMTFKLPKRAKATGWMNPQ 733

RESULT 4

COA2_PAVPN STANDARD; PRT; 749 AA.

AC P22964;
 RX MEDLINE=90010964; PubMed=2794971;
 RA Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;
 RT "Porcine parvovirus: DNA sequence and genome organization.";
 RL J. Gen. Virol. 70:2541-2553 (1989).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC -1- OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL; D00623; BAA00502.1; -;
 DR PIR; B33302; VCPVPP.
 DR HSSP; P30129; ADPV.

DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 749
 FT CHAIN 171 749
 FT DOMAIN 195 210
 FT CARBOHYD 192 192
 FT CARBOHYD 218 218
 FT CARBOHYD 302 302
 FT CARBOHYD 350 350
 FT CARBOHYD 453 453
 FT CARBOHYD 491 491
 FT CARBOHYD 593 593
 FT CARBOHYD 624 624
 FT CARBOHYD 671 671
 SQ SEQUENCE 749 AA; 82872 MW; 6CA13CB97BE93418 CRC64;

Query Match 8.1%; Score 235.5; DB 1; Length 749;
 Best Local Similarity 23.2%; Pred. No. 2.3e-09;
 Matches 139; Conservative 83; Mismatches 241; Indels 137; Gaps 32;

QY 3 SGGGAMADNNEGADVGNASGNW--CDSTWIGD--RVITSTRTWALPTYNNHLYKQ 57
 DB 194 SGGGGG--GGGGRGAGGVSTGSPFNNQTEFOYLGELVRIITAHASRLIHLNPEHETYKR 252
 QY 58 I-----SSASTGAS--NDNHFGYSTPMGYFDFNRFHCHSPBDWQRLINNNGFRPKRLN 111
 DB 253 IHLVNSGSGAGQVODDHTQWVTWMSLIDANAWGVNRPADWQILINMTIEINLVSE 312
 QY 112 FKLNIQYKEVT--TNDVTTIANNLTSTVOVFSDEYQLPVYLSAHOGLCPPPADV 168
 DB 313 QMIFNVLTITSAISPTPKIYNDLTASLWALDNTNTLPTTPAARSETLTGFPMWP 372
 QY 169 FMIPQYGL--TLNNGQAVRS-----SPYCLE-YRPSQMLRTGNNF 208
 DB 373 TKETQRYVLSCTRLNLPPTYTGQSIDTSDIGLHSDIMEFTINAVPIHLRTGDER 432
 QY 209 -TFSTFESEVPFHSVAHSQSLDR-----LNNPLI--DOVLYLNTQNOGSAQNK 257
 DB 433 STGIYHFDTKPL--KLTHSMQTRSLGLPPKLTETPTTSGDQHPGLTPAANTKKGHYIT 490
 QY 258 DILFSRGSAPAGMSVCPKMLPGPCYRQQRVSKTKTNNNSNFTWGTASKYKYLNGRESIIN 317
 DB 491 NNSYTEAT--AIR-----AQGVNTPYNNFEYSNGSP-----LT 524
 QY 318 P--GTAMASHKODEKFFPMGSMVIFGKESAGASNTALD--NVMITDEBEIK--ATNP 369
 DB 525 PIVPTADTQYNDDEPN-----GAIRFTMDYQHGHLTSSQELERYTFENPQ 569
 QY 370 -----VATERFGTV--VNFQSSS-----TDPATG--DVHAMGALP-----GMVMDQDRV 410
 DB 570 SKCGRAPKQOPNOQAPLNTENTNNTGLTLPSPDIPGGSNNHFMNTLTGPTALNNTAPV 629
 QY 411 YLOGPIWAKIPHTD--GHEFHSPLMGFGGLK--NPPOILIKTPVPAVNPAPAEFSA--TKFA 466
 DB 630 FRPGQIMDELDLIDKLRH--VTAFFQKNNPQQLRVKAP--NLTDDEPNASQPO 683
 QY 467 SFTYQSTGVSEIEMEL--OKENSKRWPEVOYTSYAKSANDFTVDNNGLYTERPPIG 526
 DB 684 PRITYSNFMWKGTLFTAKRSMNNMNIQCHTT-----TAENIGKYI--PTNIG 732

RESULT 5

COA1_PAVP9 STANDARD; PRT; 729 AA.

AC P33484;
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 (Contains: Coat protein VP2).
 OS Porcine parvovirus (strain 90HS) (PPV).
 CC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=33725;

[1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=89319168; PubMed=2750278;
 RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;
 RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";
 RL Virus Res. 13:79-86(1989).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 DR PIR; A60006; A60006.
 DR HSSP; P30129; ADPV.
 DR InterPro: IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein; Glycoprotein.
 FT CHAIN 1 729
 FT CARBOHYD 151 729 COAT PROTEIN VP1.
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 729 AA; 80938 MW; 92538BFF9A0C7886 CRC64;

Query Match 8.0%; Score 232.5; DB 1; Length 729;
 Best Local Similarity 23.0%; Pred. No. 3.6e-09;
 Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

QY 3 SGGGAPADNNEGADGVNAGSG--CDSTWLGD---RVYTTSTRTWALPTYNNHLYKQ 57
 DB 174 SGGGGG-GGGGRAGGVSTGSPNNOTEFQYIGBLVIRTAHSLIHLNMPHEHYTKR 232
 QY 58 IS--SASTGAS---NNHFGYSTPGYFDFNRFCHGSPRDQRLINNNWGRPRKLN 111
 DB 233 IHTVNSSGVAGGVQVODAHQMTPTMSLIDANMGWENADQOLISNNTEINLVSE 292
 QY 112 FKLNIQKVEY--TNDGVTTANNITSTVOVFSDEYOLPYVLSAHOGLPPPADV 168
 DB 293 QELFNVLKTTITESATSPPTKIYNNDLTASLMVALLDTNNTLTPYAPARSETLGFYPLP 352
 QY 169 FMIPQYVL-----TLNNSQAVGRS-----SFYGLF-YFSPQMRTGNNF 208
 DB 353 TKPTQYRYVLSCTRNINPPTYTGSQOITDSIQGLHSDIMEFYTENAVPIHLRTGDEF 412
 QY 209 -TESYTFEEVDFHSSVYASGSLDR-----LNNPLI--DCVLYYNNRTONOGSAQNK 257
 DB 413 STGIYHDTKPL--KLTHSQWNTNSLGLPRKLLEPTTEBQGHGTLPAANTRGHQYQM 470
 QY 258 DLLFSRGSPPAGMSVQPKWMLPGCYRQORVSKYTTDNNNSNFTVTAASKYNLNGRESIIN 317
 DB 471 NNSYTEAT---AIRP-----AQGVYNNPYNMFESNGGPF-----LT 504
 QY 318 P--GTAAASHKDDKRFPMGKSGVWIFGKESGASNTALDNWITDESEIKATNP----- 369
 DB 505 PIYPTADTQYNDSD---PNCALFTWGYOHGULTTS---SGLERYTFENQSAKGR 554
 QY 370 WATERFETVA-VNFQSS-----TDPATG--DVHANGALP-----GMWQDRDVLVQGP 415
 DB 555 APKQGFQOQAPLNLENNTNNGTLPSDPIGCKPMHFNNTLNTYGLPTALANTAVFPNGQ 614
 QY 416 IWAKIPTTD--GHFSPRLMGGFGLK-NRPQILIKNTVPVAPNPAEPA-TKFASTITQ 471
 DB 615 IMDKELDTDKPRHL---VTAPFVCKNNPFGQLFVKIAP---NLTDENADSPOQPRIT 668
 QY 472 YSTGQVVEIEEMELQKENSKRWNEVOYTSVYASAVNDFTVONGGLYTERPRIG 526
 DB 669 YSNFMWKGTLFTAKMRSSNNMNDIQOHTT-----TLENIGNYI-PTNIG 712

RESULT 6

COAL_PAVEN STANDARD; PRT; 729 AA.

AC P18546; Q89816;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein vp1 [contains: Coat protein VP2] (Version 1).
 OS Porcine parvovirus (strain NADL-2) (PPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=10797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021005; PubMed=2219713;
 RA Vasudevaracharya J., Basak S., Srinivas R.V., Compans R.W.;
 RT "The complete nucleotide sequence of an infectious clone of porcine
 RT parvovirus, strain NADL-2.";
 RL Virology 178:611-616(1990).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9085785; PubMed=2596019;
 RA Vasudevaracharya J., Basak S., Srinivas R.V., Compans R.W.;
 RT "Nucleotide sequence analysis of the capsid genes and the right-hand
 RT terminal palindromic of porcine parvovirus, strain NADL-2.";
 RL Virology 173:368-377(1989).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M38367; AAA46919.1; -;
 DR EMBL; M38367; AAA46921.1; -;
 DR EMBL; M32787; AAA46917.1; -;
 DR EMBL; M32787; AAA46918.1; -;
 DR PIR; B33743; VCPVNA.
 DR HSSP; P30129; ADPV.
 DR InterPro: IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein; Glycoprotein.

FT CHAIN 1 729
 FT CARBOHYD 151 729 COAT PROTEIN VP1.
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 175 190 GLY-RICH.
 SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80D842 CRC64;

Query Match 8.0%; Score 232.5; DB 1; Length 729;
 Best Local Similarity 23.0%; Pred. No. 3.6e-09;
 Matches 138; Conservative 85; Mismatches 240; Indels 137; Gaps 32;

QY 3 SGGGAPADNNEGADGVNAGSG--NNHCDSTWLGD---RVYTTSTRTWALPTYNNHLYKQ 57
 DB 174 SGGGGG-GGGGRAGGVSTGTFNNOTEFQYIGBLVIRTAHSLIHLNMPHEHYTKR 232
 QY 58 IS--SASTGAS---NDNHFGYSTPGYFDFNRFCHGSPRDQRLINNNWGRPRKLN 111

Db 233 IHWLNSESGVAGQVODDHAHTQVTPMSLIDANAMGVFNPAWQOLISNNMTEINLVSE 292
 QY 112 FKLENIQVKEVT---TNDGVTITANNLTSTVOVPSDEYOLPYVLSAHQGLPPPPADV 168
 Db 293 QEIFNVVLKTTTESATSPPTKIYNNDLTASLWALDNTNLTPTPAAPSEETLGFPWLP 352
 QY 169 FMIPQYGL-----TLNNGSOAVGRS-----SFYCLE-YFPQOMLRGNF 208
 Db 353 TKPTQRYVLSCTRNANPPTVYGSOQITDSIQGLHSDIMFYTIEANVAIVHILRTGDEF 412
 QY 209 -TFSTYEEVPPHSSVYHQSGLR-----LMNPLI--DOYLYYLNFTQNGSGAQNK 257
 Db 413 STGIYHDTKPL--KLTHSMQWTRSLGLPPKLLTEPTTEGQHOGHTLPAANTRGYHQT 470
 QY 258 DLLFSRGSFAGMSVQPKMWLPGPCYRQGRVSKTIDNNNSNFTWTSKYNLNGRESIIN 317
 Db 471 NNSYTEAT---AIRP-----AOGVYTPYNNFEYSNGGPP-----LT 504
 QY 318 P--GTAMASHKDEDEKFFPMGSGWIFGKESAGASNTALD---NVMITDEEIK--ATNP- 369
 Db 505 PIYPTADTQYNDENP-----GAIKFTMDYOHGLITSSQELERYTFNPQ 549
 QY 370 ----VATERFGTVA-VNFOSS-----TDPATG--DVHMGALP-----GMVWQDRDY 410
 Db 550 SKCGRAPKQOFNOQAPLNTENTNNTLLPSDPIGKSNMFMNTLNTYGLTALNTAPV 609
 QY 411 YLOGPIWAKIPTHD--GHFHSPLMGFGGLK--NPPQILIKNTVPANPPAEBSA--TKFA 466
 Db 610 FPNQQLMDKLDLTKPRHL--VTAPFVCKNPPGQLFVKIAP--NLTDDFNADSPQ 663
 QY 467 SFTQVSTGVSEIEMELQKENSRRNPEVQYTSNKAASNVFTVDDNGLYTEPRPIG 526
 Db 664 PRITTSNFMWKGTLTFTAKRSSNMNPIQOHTT-----TANIGNYI-PTNIG 712

RESULT 7

COAT_PAVPK STANDARD; PRT; 729 AA.
 ID AC P52501;
 DT 01-OCT-1996 (Rel. 34, Last Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2].
 OS Porcine parvovirus (strain Kresne) (PPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NC NCB1_TaxID=73487;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE=96183900; PubMed=8642680;
 RA Bergeron J., Hebert B., Tjissen P.;
 RT "Genome organization of the Kresne strain of porcine parvovirus:
 RT identification of the allotropic determinant and comparison with
 RT those of NADL-2 and field isolates.";
 RT J. Virol. 70:2508-2515(1996).
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL databank on its
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U44978; AAC40230.1; -;
 DR EMBL: U44978; AAC40231.1; -;
 DR HSSP: P30129; ADPV.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.

KM Coat protein; Glycoprotein. COAT PROTEIN VP1.
 FT CHAIN 1 729
 FT CHAIN 151 729
 FT CARBOHYD 172 172
 FT CARBOHYD 198 198
 FT CARBOHYD 282 282
 FT CARBOHYD 330 330
 FT CARBOHYD 433 433
 FT CARBOHYD 471 471
 FT CARBOHYD 573 573
 FT CARBOHYD 604 604
 FT CARBOHYD 651 651
 SQ SEQUENCE 729 AA; 80835 MW; B6345BBA0568A1F6 CRC64;

Query Match 7.9%; Score 230.5; DB 1; Length 729;
 Best Local Similarity 23.0%; Pred. No. 5e-09;
 Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

QY 3 SGGGAEMADNNEGADVGNAAGNMH--CDSTWLGD---RVITTSRTMALPTYNHLYKQ 57
 Db 174 SGGGGG-GGGGRAGGVSTGSPFNNOTFQYIGELVATITAHASLILHLMPEHETYGR 232
 QY 58 IS--SASTGAS---NDNHYFGYSTFWGYFDNRFPHCHSPRDWQRLINNNMGFRKRLN 111
 Db 233 IHWLNSESGVAGQVODDHAHTQVTPMSLIDANAMGVFNPAWQOLISNNMTEINLVSE 292
 QY 112 FKLENIQVKEVT---TNDGVTITANNLTSTVOVPSDEYOLPYVLSAHQGLPPPPADV 168
 Db 293 QEIFNVVLKTTTESATSPPTKIYNNDLTASLWALDNTNLTPTPAAPSEETLGFPWLP 352
 QY 169 FMIPQYGL-----TLNNGSOAVGRS-----SFYCLE-YFPQOMLRGNF 208
 Db 353 TKPTQRYVLSCTRNANPPTVYGSOQITDSIQGLHSDIMFYTIEANVAIVHILRTGDEF 412
 QY 209 -TFSTYEEVPPHSSVYHQSGLR-----LMNPLI--DOYLYYLNFTQNGSGAQNK 257
 Db 413 STGIYHDTKPL--KLTHSMQWTRSLGLPPKLLTEPTTEGQHOGHTLPAANTRGYHQT 470
 QY 258 DLLFSRGSFAGMSVQPKMWLPGPCYRQGRVSKTIDNNNSNFTWTSKYNLNGRESIIN 317
 Db 471 NNSYTEAT---AIRP-----AOGVYTPYNNFEYSNGGPP-----LT 504
 QY 318 P--GTAMASHKDEDEKFFPMGSGWIFGKESAGASNTALNVMITDEEIKATNP----- 369
 Db 505 PIYPTADTQYNDENP-----GAIKFTMDYOHGLITSSQELERYTFNPQ 549
 QY 370 VATERFGTVA-VNFOSS-----TDPATG--DVHMGALP-----GMVWQDRDYLOGP 415
 Db 555 APRQGFNOQAPLNTENTNNTLLPSDPIGKSNMFMNTLNTYGLTALNTAPVFPNGQ 614
 QY 416 IMAKIHTD--GHFHSPLMGFGGLK--NPPQILIKNTVPANPPAEBSA--TKFASFITQ 471
 Db 615 IMDKELDTLTKPRHL--VTAPFVCKNPPGQLFVKIAP--NLTDDFNADSPQPRIT 668
 QY 472 YSTQVSTGVSEIEMELQKENSRRNPEVQYTSNKAASNVFTVDDNGLYTEPRPIG 526
 Db 669 YSNFMWKGTLTFTAKRSSNMNPIQOHTT-----TANIGNYI-PTNIG 712

RESULT 8

COAT_PAVL3 STANDARD; PRT; 587 AA.
 ID AC P36310;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2].
 OS Parvovirus lu111.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NC NCB1_TaxID=35339;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE=93297126; PubMed=8517025;

RA Diffoc N., Chen K.C., Bates R.C., Lederma M.;
 RT "The complete nucleotide sequence of parvovirus h1u1 and
 RT localization of a unique sequence possibly responsible for its
 RT encapsidation pattern."
 RL Virology 192:339-345 (1993).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M81888; NOT_ANNOTATED_CDS.
 DR PIR, B44276; B44276.
 DR HSSP, P07302; IMVM.
 DR InterPro, IPR001403; Parvo_coat.
 DR Pfam, PF00740; Parvo_coat.1.
 DR Cost protein, Glycoprotein.
 KM CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 587 AA; 65429 MW; 523E6B9CBF2EBB74 CRC64;

Query Match 7.9%; Score 229; DB 1; Length 587;
 Best Local Similarity 22.7%; Pred. No. 4.8e-09;
 Matches 133; Conservative 74; Mismatches 228; Indels 150; Gaps 28;

3 SGGGAPMADNNEGADGVGNAGNMHCDS--TWLGD--RVITTSRTWAL--PTYN-- 52
 30 SGGG-----GGGGGVGVSTGSDYNQTHYKFLDGMWEITNYSTRWHLNMPSENCR 83
 53 --HLKYGISASTGASNDNHYFGYSTPWGYFDENRFCHGSPRDQRLINNMGFRPKR 109
 84 VRVHNNTDGTASHAMDAHQIWI--TPMSLVDAAMGVWFOPSDQWYISNMHILNLS 142
 110 LNFKLFIQVKEVT--TNDGVTTIA--NNLSTGVYEDSEYOLPYVLGSAHQGL-- 161
 143 LDOELFNVIKIVTEQNTGAEAIKYNNDLTAAAMVVALDSNNILPYTPALDNOETLGFYP 202
 162 --PPPPA-----DVFMLPQYGLTLNNGSAVGRSSFYCLEYFPPS--QMLRTGN 206
 203 WKPTPSYRYRYFSCDRRLSTLYKDEAGTITDTMGASGLNSQPTTIENTQRIHLKRGD 262
 207 NF-TFSYTFEEVPHFSSYAHQSQSLRLMNPILDQ-----YLYINRTGN- 249
 263 EVATGYTFEDIEPIRLTLHTWQNRHLGQPPQTELPSSDTANATLTARGYSGLTQIGR 322
 250 -----QSSSAQNKD--LFSRSGPAGMSYQPKMLPG-----PCYRQR 286
 323 NDVTEATVRPAQVGFQPHDNFETSRAGPFKVPVAPADITQGLDHDANGSLRYTYDKOH 382
 287 VSKTKDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESA 346
 383 GQSMASQNKRYRW-DAVNTD--SGR----- 406
 347 GASNTALDNMTDEEIKATNPATER--FGTVAANFQGS--SSTDATGDIAMGALP 401
 407 WTNNGFIQSVFPTSBN--ANQILTNRDNLAKGDIHFTNAFNSYGLTAFPH----- 457
 402 GAWMDRQVYLOGPIYAK--IPHTDGHFHSPLMGSGGLKNPPQILIKRTP--VPANP 456
 458 -----PADIYQGGQIWDKELDEHKPRILHTQAQPV--CKNNAPQILLVRLAENLTDQYD 509

457 PAEFATKPAFTIYOSTGVSEIEMELQKENSKEWNEVQYTS 501
 510 PMSNLSRIYTYGTFPMKGLTIKAK---MRPNA-TWNPVFOISA 550

RESULT 9
 COAT_PAVNH STANDARD; PRT; 722 AA.
 ID COAT_PAVNH
 AC P03136;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 [Contains: Coat protein VP2].
 OS Hamster parvovirus H1.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 OX NCBI_TaxID=10799;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8312183; PubMed=6823009;
 RA Rhode S.L. III, Paradiso P.R.;
 RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its
 RT genes by hybrid-arrested translation."
 RL J. Virol. 45:173-184 (1983).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X01457; CAB57285.1; ALT_SEQ.
 DR PIR, A03699; VCPV2.
 DR HSSP, P07302; IMVM.
 DR InterPro, IPR001403; Parvo_coat.
 DR Pfam, PF00740; Parvo_coat.1.
 KM Coat protein, Glycoprotein.
 FT CHAIN 1 722
 FT CARBOHYD 131 722 COAT PROTEIN VP1.
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 647 647 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 155 170 GLY-RICH.
 SO SEQUENCE 722 AA; 79737 MW; 6BB678391AASDC31 CRC64;

Query Match 7.8%; Score 226; DB 1; Length 722;
 Best Local Similarity 21.7%; Pred. No. 1e-08;
 Matches 128; Conservative 83; Mismatches 234; Indels 146; Gaps 26;

2 ASGGAPMADNNEGADGVGNAGNMHCDS--LGDRTVTSRTWAL----- 47
 153 ADGGG--SGGGGGGGGIGVSTGYNDQTYKFLDGMWEITASHRLHGMPPSENYC 211
 48 --PTYNHLYKYGISASTGASNDNHYFGYSTPWGYFDENRFCHGSPRDQRLINNMWGF 105
 212 RVTVHNNQTTGHGKVKGNAPYDTHQIWI--TPMSLVDAAMGVWFOPSDQWYISNMH 270
 106 RPKLNFKLFIQVKEVT-----NDGVTTIANLSTVOVFSDEYOLPYVLGSAHQGC 160
 271 NLSLSGELFNVIKIVTEQNTGAEAIKYNNDLTAAAMVVALDSNNILPYTPAQTSRT 330
 161 L-----PPPPAD--VFMIPQYGLTLNNGSQ-----AVG-----RSSFYCLE-YFPPS 199
 331 LGFPMPTAPAPARYRYFFMPROLSTVSSNSAGTQITDTIGEPQALNSQPTTIENTLPI 390
 200 QMLRTGNFPP--SYTFEEVPHFSSYA-----HQSQSLRLMNP-PLIDQVLYLNRTONQSG 252
 391 TLRTGDEFTTYGTFIYNTDPLKLHTWQNRHLAQLGQITDLPSTATATSLANGDRFG 450

QY 253 SAONKLLF-----SRGSPAGNSVOPKMWLPGPCYRQGRVSK 289
 Db 451 STQGNVNYVTEALRTRPAQIGFWQPHDNPEANGGFKVPVP-----LDI 497
 QY 290 TKTDNNNSNTWGTGASKYNLNGR--ESI INPGLT-----MASHKDEDEKFFPMG 337
 Db 498 TAEDEHDHAN-----GAIRFNYSKQGHGEWMAKGAAPERYTWDALDASAGRTARCFV-----549
 QY 338 VMTFGKSGAGSANTALDNWMTDEEELKA--TNPATERFGTVANFGSSSTDPATGVHA 396
 Db 550 -----OSAPISIPPNQOQLQREDALAGRTMHTVFNVSYPGLSAFPHBDP-----596
 QY 397 MGALPGWQDRDVLVLOGPIWAK--IPHTDGHFHPSPLMGFGGLKNPPOQILIKNTPPV 453
 Db 597 -----IYPMGOIMDKELDEHKRPLHTVAPFV-----CKNNPPQGLFVHLGP-- 638
 QY 454 ANPPAEF--SATKFASTTQYSTGQVSVEIEMELQENSKRNPEVOYTSN 502
 Db 639 -NLTDQDPNSTVTSRIVT-YSTFYWKGLIKFKAKLRLNLTWNPVQATTD 687

RESULT 10

COAT_MUMIM STANDARD; PRT; 716 AA.

AC PO3137;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 (Contains: Coat protein VP2).
 OS Murine minute virus (Murine parvovirus).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=10794;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83143341; PubMed=6298737;
 RA Astell C.R., Thomson M., Merchinsky M., Ward D.C.;
 RT "The complete DNA sequence of minute virus of mice, an autonomous
 RT parvovirus."
 RL Nucleic Acids Res. 11:999-1018 (1983).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: V01115; CAA24310.1; ALT_SEQ.
 DR PIR: A03700; VCPV2M.
 DR HSSP: P07302; 1MVM.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 / 716 COAT PROTEIN VP1.
 FT CARBOHYD 179 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 156 171 GLY-RICH.
 SQ SEQUENCE 716 AA; 78707 MW; 6A7229A91161F4C6 CRC64;

Query Match

Best Local Similarity 6.8%; Score 197; DB 1; Length 716;
 Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;

QY 3 SGGAPADNNEGADGYNAGSNHCHS--TWLGD---RYTTSTRTWALPTYNHLYKQ 57
 Db 160 SGGG-----GSGGGGVSTGYSYDNTHYRFLDGDGWEITATRLVHLMPSSENYCR 213

QY 58 I-----SSASTGASNDNHYFGYSTPMGYFDFNRFCHFSPPRQRLINNMGFRPRKL 110
 Db 214 IRHANTDPSVKNMNAKDABEOITPMSGLVDNANAGVWLQSPDWQYICNTMSQLNLVSL 273
 QY 111 NFKLENIQVEKVTND---GVTIANLSTVQVFSDBYQLPYVLSAHQGLPPFPA 166
 Db 274 DQEIFNVVAKTVTEODLGGQAIKIYNNDDLACMMVAVDSNNILPYTPANSMETLGFYWP 333
 QY 167 DVFMIOQYQ-----LTNNGSOAV-----GRSSPYCLEYFSS--OMLSTGN 207
 Db 334 KPTIASPYRYFCVVDRLSTYENQGTVEHNMVGPKGIPQPFTEHTQOITLLRTGDE 393
 QY 208 F-TFSYTFEEVPRHSSVYASQSLRLMN--PLIDQYLYLNRTONSGSARONKDLLFSRG 264
 Db 394 FATGYTFPDNSV--KLTHTWQNRQLGQRPILSTF-----PEADTQAGT-----LTAQ 441
 QY 265 SPAGMSVOPKMWL-----PGPCYRQGRVSKRTDNNNSFT 300
 Db 442 SRHGTQMGVNWVSEAIRTRPAQVGFCCPHNDEEASRAG-FAAPKVPADITQGVDEKAN 500
 QY 301 WTGASKYNLNGRESIINPG-----TAMASHKDEDEKFF-----PMGVMIF 341
 Db 501 GSVRYSGKHQGHENMASHGAPERYTWDETSFGSGDPTQDGTQSAPLVVPPLNGI-- 557
 QY 342 GKESAGASNTALDNWMTDEEELKATNPATERFGTVAVNFQ--SSTDPATGDVHAMGA 399
 Db 558 -----LTNANPIGTGN--DIHFSVFNYSYGLTAFSH-----587
 QY 400 LPMGWQDRDVLVLOGPIWAK--IPHTDGHFHPSPLMGFGGLKNPPOQILIKNTP--VPA 454
 Db 588 -----PSPVYPOQGIWMDKELDEHKRPLHTVAPFV-----CKNNAPGQMLVRLGPNLTQ 637
 QY 455 NPPEAFSATKFASTTQYSTGQVSVEIEMELQENSKRNPEVOYTSNAXKSNVPTVD 514
 Db 638 YDNGATLSRIYVYGFPMWKGKLTMPAKLRA--NTWNPVYQSAE-----D 682
 QY 515 NNGLY--TEPRPIGT-----RYLTRPL 534
 Db 683 NGNSVMSVTKWLPATGAGNQSVPILTRPV 711

RESULT 11

COAT_MUMIM STANDARD; PRT; 718 AA.

AC PO7302; Q9WMH3; Q9WMH2;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 (Contains: Coat protein VP2).
 OS Murine minute virus (strain MMV1) (Murine parvovirus).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI_TaxID=10795;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85242059; PubMed=3855242;
 RA Sahli R., McMaster G.K., Hirt B.;
 RT "DNA sequence comparison between two tissue-specific variants of the
 RT autonomous parvovirus, minute virus of mice."
 RL Nucleic Acids Res. 13:3617-3633 (1985).
 [2]
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.
 RA Llamas-Saiz A.L., Agbandje-Mckenna M., Wikoff W.R., Bratton J.,
 RA Tattersall P., Rosemann M.G.;
 RT "Structure determination of Minute Virus of mice."
 RL Acta Crystallogr. D 53:93-100 (1997).
 [3]

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02481; CAB46507.1; -
CC EMBL; X02481; CAB46508.1; -
CC EMBL; M12032; AAA69569.1; ALT_INIT.
CC PIR; B23008; VCPVIM.
CC DR; 1MVM; 25-FEB-98.
CC DR; InterPro; IPR001403; Parvo_coat.
CC DR; PIR; PF00740; Parvo_coat; 1.
CC KW; Coat protein; Glycoprotein; 3D-structure.
CC FT; CHAIN 1 718 COAT PROTEIN VP1.
CC FT; CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 502 502 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; DOMAIN 157 172 GLY-RICH.
CC FT; COMPLET 144 144 A -> G (IN REF. 2).
CC SO; SEQUENCE 718 AA; 79017 MW; B43C1762ED5F4B9 CRC64;
Query Match 6.8%; Score 196.5; DB 1; Length 718;
Best Local Similarity 20.5%; Pred. No. 1.4e-06;
Matches 122; Conservative 82; Mismatches 243; Indels 147; Gaps 25;
QY 3 SGGGAPADNNEGADGVGNAGSNHCHDS--TWLGP--RVITTSRTALPTNNHLYKQ 57
DB 161 SGGG-----SGGGGAGVSTGSGSYNQTHRPLGAGWEITATLRLVHLMPSKENC 214
QY 58 I-----SSASTGASNNHFGYSTPMGVPDFNRFCHFSRDPORLNNMGFRPRKL 110
DB 215 IRYNHTTTSYKGNNAKDAHQITWTPSLVDANAGWLPDSMOYICNTMSQNLVSL 274
QY 111 NFKLFNIQVKEVTND---GVTTIANLJSTVOYFSDSEYQLPVYLGSAHQCLPPPPA 166
DB 275 DOEIFNVVLTVTBEDSGQAQIKIYNNDLTACMMVAVSNMILPTPAANSMETLGFYFW 334
QY 167 DVEMIPQVY-----LTLAN-----GSAQVRSFPCLEFPFS-QMLRTGN 206
DB 335 KPTIASPRRYPCVDRLDLSVTYENQGTIEHNVMGTFRGNNSQFTIENTQOITLRTGD 394
QY 207 NF-TFSYTFEEVFPFSSSYAHSQSILRLNN--PLIDQVLYLIRTONOGSANKDLFSR 263
DB 395 EFATGTYTFDNPV--KLTHWQNRKOLGPPILSTF-----PEADTDAGT-----LTAQ 442
QY 264 GSPAGMSVYQPKWL-----PGRCTYRQQRVSKTKTNNNSNF 299
DB 443 GSRGATQMEVNWVSEAIRTRPAQVFCQPHNDFASRAGP-FAAPKPAVDVTCVDBEA 501
QY 300 TWLGASKNLNGREBINPTAMASHKDEDF-----FPMSSGMITGKSAGASN 350
DB 502 NGSVRSYSGKHGEMWAHAGPAPERITYWDETNPFGSDTRDGFIOASAPLVVPPPLNGILT 561
QY 351 TALNDVMTIDREELIKATNPVATERGTAANFOS--SSTDPAATGVHAMGALPGWVQDR 408
DB 562 NA-----NPIGTKN-----DIHFSNVFNISYGLPIAFSH-----PS 591
QY 409 DVYLGGPIWAK--IPHTDGHFHSPLMGFGGLKXNPPQILIKTP--VPANPAEBSAT 463
DB 592 FVYVPGQIWDKRLDLHKKRPLHITAPFV--CKNNAFGOMLVRLGPNLTDQYDPRGATLS 648
QY 464 KFASTITQYSTGQSVLEIEMELQKENSGRANPEVOYTSNVAKSANVDPTVONNG 517
DB 649 RIVYVGTFFWKGKLTMRAKLRA-----NTTNPVVO-----VSVEDNG 686

RESULT 12
ID COAT MEVA STANDARD; PRT; 722 AA.
AC P27437;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2].
OS Mink enteritis virus (strain Abashiri) (MEV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
RX NCBI_TaxId=10793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9120123; PubMed=2016597;
RA Kariatsunari T., Horuchi M., Hama E., Yaguchi K., Ishiguro N.,
RA Goto H., Shinagawa M.;
RT "Construction and nucleotide sequence analysis of an infectious DNA
RT clone of the autonomous parvovirus, mink enteritis virus";
RT J. Gen. Virol. 72:867-875(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00765; BAA00663.1; -
CC PIR; B38350; VCPVME.
CC DR; HSSP; P31029; ADPV.
CC DR; InterPro; IPR001403; Parvo_coat.
CC DR; PIR; PF00740; Parvo_coat; 1.
CC KW; Coat protein; Glycoprotein.
CC FT; CHAIN 1 722 COAT PROTEIN VP1.
CC FT; DOMAIN 139 722 GLY-RICH.
CC FT; CARBOHYD 160 177 GLY-RICH.
CC FT; CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT; CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SO; SEQUENCE 722 AA; 79823 MW; 9DADBCAB2EF9F622 CRC64;
Query Match 6.7%; Score 194; DB 1; Length 722;
Best Local Similarity 20.7%; Pred. No. 2.1e-06;
Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;
QY 5 GGAAPADNNEGADGVGNAGSNHCHDSITWLGDRYITTS-----RTALPTNN- 52
DB 148 GGQP-AVRNERATGSGGSGGGGGS--GGVGISGTENNQTEFKFLNGWEITANSS 204
QY 53 -----YKQI-----SSASTGASNNHFGYSTPMGVPDFNRFCHFSRPRW 95
DB 205 RLVHLMPSSENYKRYVANNMDKTAVKGNALDDTHVQIYTPMSLYDANAGWTFNPGDW 264
QY 96 QRLINNNMGFRPRRLNFKLFNIQVKEVT--TNDGVTTIANLJSTVOYFSDSEYQLPVY 152
DB 265 QLVNTMSSELHVSFPGIEIFNVVLTKVYSESAQTPPKRVYNNDLTASLMDVLDNNTMPFT 324
QY 153 LGSNAGCL-----PPFPA-----DVMTLPQYGLTLNNGSAQVRS----- 189
DB 325 PAAMRSETLGFYFWKCTIIPPMRYTFQWDRTLIPSH-----TGTSGTPNIVHGTPDD 378
QY 190 -SFYCLE-YFPQOMLRTGNVF--TF-----SYTFEE-----VPHSSYASQSL 229
DB 379 VQFTTENSVPVHLRTIGDEFATGTFPPDCKPRLTHWQNRALGLPPLNLSPOSSEGA 438
QY 230 DR-----LMNPLIDQYL--YYLNRTONOS----- 251

Db 439 TNFGDIGVQDKRGVYQMGNTDYITEATIMRPAVGYSAAPYGSFASTQGFPIAAG 498
 Qy 252 -GSAQNKDILFSRSGPAGMSVQPKMLPGPCY---RQQRVSKTKTDNNNSNFTWTGASKY 307
 Db 499 RGGAGTDEMOAAGDPR-----YAFGRHQGOKTTTGTETPERFTY----- 538
 Qy 308 NLNGRESIINGTAMASHKDEDEKFFPMGSGVIMFGKESAGASNTALD-NVMITDEEIK 366
 Db 539 -----IAHQDT-----GRYPAGDMQININPILPTINDVLLP 570
 Qy 367 TNPVATERFGTVAVNFQSSSTDPAQDVHAMGALPGWVQDRDYLQGPIMAKIPHTDGH 426
 Db 571 TDPFG-----GKTGINY--TNIFNTYGPLTALNNV-----PYVPNGQIMDKPEPTD-- 615
 Qy 427 FHSPPLMGGRGK-----NPPQILIKTTPVPAAN---PPAESATKFSFPT 470
 Db 616 -----LKRRLHVNAPFVCQNNCPQLFVKVAPNLITNEYDPASANMSR-----IV 660
 Qy 471 QYSTGVSEIEMELQKNSKRNPEVOYTSNVAKSANVDPTDNNGLYTEPRPIG 526
 Db 661 TYSDFPMWKGVLVFKAKLRASHHTNPICQMSIN-----VDNQPNYL-PNNIG 705

RESULT 13
 ID COAT_PPV STANDARD; PRT: 727 AA.

AC P04864; 065112;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coat protein VP1 (Contains: Coat protein VP2).
 OS Feline parvovirus.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI TaxID=10786;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85265017; PubMed=2991581;
 RA Carlson J., Rushlow K., Maxwell I., Maxwell F., Winston S., Hahn W.;
 RT "Cloning and sequence of DNA encoding structural proteins of the
 RT autonomous parvovirus feline parvovirus.";
 RL J. Virol. 55:574-587(1985).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

CC -----
 CC EMBL; M10824; AAA47161.1; -;
 DR EMBL; M10824; AAA47162.1; -;
 DR PIR; A03701; VCPVIF.
 DR HSSP; P30129; ADPV.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 727
 FT CHAIN 1 727
 FT CARBOHYD 144 727 COAT PROTEIN VP1.
 FT CARBOHYD 168 168 COAT PROTEIN VP2.
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DOMAIN 165 180 GLY-RICH.
 SQ SEQUENCE 727 AA; 80344 MW; 77E0FEF554C0C6E CRC64;

Query Match 6.7%; Score 194; DB 1; Length 727;
 Best Local Similarity 20.7%; Pred. No. 2.1e-06;

Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;

Qy 5 GGAADNNEGAGVGNASGNHCDSTWIGDRITST-----RTALPTVNN- 52
 Db 153 GQGP-AVRNERAAGSNAGGGGGG--GGVIGSTGNNQTEKFLENGWEITANSS 209
 Qy 53 ---HL-----YKQI-----SSASTASNDNHFGSTPWGDFNRFCHGFSRDM 95
 Db 210 RLVLHLMPSSENKRVVANNMDKTAVKGNMALLDTHQVITPMSLVDAAMGWGFRGDK 269
 Qy 96 QRLINNNWGPFRRLNFKLNIQVKEVT--INDGVTTIANNTSTVQVPSSEYQLPVY 152
 Db 270 QLIVNTMSELHLVSPFOELFNVVLKTVSSATQPPKVVNNDLTASLMAVLDNNTPMPT 329
 Qy 153 LGSAGGCL-----PPFPA-----DVPMIPOGYLILNNGSAVRS----- 189
 Db 330 PAAMRETTIGFYWKRTLPFWRYFQWBRLLPSH-----TGTGTPNTIYHGTPDD 383
 Qy 190 -SFYCLE-YFPQMLRTGNF--TF-----SYTEE-----VPFSSVYASQSL 229
 Db 384 VQFTTENSVPVHLATGDEFATGTFPFDCKPCLRLHTQTNALGLPPLNLSPOSEGA 443
 Qy 230 DR-----LMDLIDQYL--YLLNFTQNS----- 251
 Db 444 TNFGDIGVQDKRGVYQMGNTDYITEATIMRPAVGYSAAPYGSFASTQGFPIAAG 503
 Qy 252 -GSAQNKDILFSRSGPAGMSVQPKMLPGPCY---RQQRVSKTKTDNNNSNFTWTGASKY 307
 Db 504 RGGAGTDEMOAAGDPR-----YAFGRHQGOKTTTGTETPERFTY----- 543
 Qy 308 NLNGRESIINGTAMASHKDEDEKFFPMGSGVIMFGKESAGASNTALD-NVMITDEEIK 366
 Db 544 -----IAHQDT-----GRYPAGDMQININPILPTINDVLLP 575
 Qy 367 TNPVATERFGTVAVNFQSSSTDPAQDVHAMGALPGWVQDRDYLQGPIMAKIPHTDGH 426
 Db 576 TDPFG-----GKTGINY--TNIFNTYGPLTALNNV-----PYVPNGQIMDKPEPTD-- 620
 Qy 427 FHSPPLMGGRGK-----NPPQILIKTTPVPAAN---PPAESATKFSFPT 470
 Db 621 -----LKRRLHVNAPFVCQNNCPQLFVKVAPNLITNEYDPASANMSR-----IV 665
 Qy 471 QYSTGVSEIEMELQKNSKRNPEVOYTSNVAKSANVDPTDNNGLYTEPRPIG 526
 Db 666 TYSDFPMWKGVLVFKAKLRASHHTNPICQMSIN-----VDNQPNYL-PNNIG 710

RESULT 14
 ID COAT_PAVC2 STANDARD; PRT: 584 AA.

AC P30129;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Coat protein VP2.
 OS Canine parvovirus (type 2 / strain A72) (CPV).
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
 NCBI TaxID=31597;
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
 RX MEDLINE=91173306; PubMed=2006420;
 RA Tao J., Chapman M.S., Agbandje M., Keller W., Smith K., Wu H.,
 RA Luo M., Smith T.J., Roseman M.G., Compans R.W., Parish C.R.;
 RT "The three-dimensional structure of canine parvovirus and its
 RT functional implications.";
 RL Science 251:1456-1464(1991).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC OR A COMBINATION OF VP2, VP3, AND SOME VP1.
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUS COAT PROTEIN FAMILY.
 CC -----
 CC PDB; 4DEV; 01-APR-97.
 DR PDB; 1JUS; 23-DEC-96.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.

KW: Coat protein; Glycoprotein; 3D-structure.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 39
FT STRAND 49 53
FT STRAND 58 64
FT STRAND 67 71
FT STRAND 80 84
FT TURN 87 89
FT STRAND 102 112
FT TURN 118 120
FT HELIX 123 132
FT STRAND 133 154
FT STRAND 164 167
FT STRAND 169 170
FT STRAND 173 177
FT TURN 179 180
FT TURN 188 191
FT TURN 199 200
FT STRAND 203 203
FT STRAND 206 211
FT STRAND 215 218
FT STRAND 231 235
FT TURN 238 239
FT HELIX 246 249
FT STRAND 252 255
FT STRAND 261 262
FT STRAND 266 267
FT STRAND 270 270
FT STRAND 273 274
FT TURN 300 301
FT HELIX 310 312
FT TURN 317 317
FT STRAND 327 329
FT STRAND 330 330
FT STRAND 335 336
FT STRAND 339 339
FT STRAND 346 347
FT STRAND 352 353
FT STRAND 363 364
FT TURN 367 368
FT TURN 373 374
FT STRAND 378 380
FT TURN 388 389
FT STRAND 396 398
FT TURN 410 411
FT STRAND 415 416
FT TURN 418 419
FT HELIX 426 428
FT STRAND 429 430
FT TURN 445 447
FT STRAND 450 450
FT TURN 453 454
FT STRAND 455 456
FT STRAND 464 464
FT TURN 466 466
FT STRAND 470 470
FT STRAND 488 488
FT STRAND 490 490
FT STRAND 498 502
FT STRAND 506 506
FT TURN 512 513
FT STRAND 520 520
FT STRAND 523 539
FT TURN 555 557
FT HELIX 558 560
FT STRAND 565 566
FT STRAND 574 574

SQ SEQUENCE 584 AA; 64686 MW; 6E4DADA5AEEF9DBC CRC64;
Query Match 6.5%; Score 190; DB 1; Length 584;
Best Local Similarity 20.3%; Pred. No. 3e-06;
Matches 130; Conservative 82; Mismatches 229; Indels 198; Gaps 30;
OY 5 GGAPMADNNGAGGVGNASGNMHCDSITWLDGRYITTS-----RTMALPTVNN- 52
10 GGCP-AVRNERAAGSGNGSGGGGGG--GGVGLSTGTFFNNQTEPKLENGWYITNNS 66
DB 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPWGYPDEFNCHFSFDDW 95
67 RLVLNMPESSENYRVRVNNMDKTAVGNMALLDIAEIVTIPMSLVANAGWYFNGDW 126
OY 96 ORLLNNWGFRRPKRINKLENIQKVT---TNDGVTIANNLITSTQVSDSEYOUPYV 152
127 QLVNTNSELHLVSFEQELFNVLKTVSESATQPPKTVNNDLTASLMLVLSNNITWFT 186
OY 153 LGSNQGCL-----PPFPA-----DVFMIPQYGYLTNNGSAVGRS----- 189
187 PAARSETLGFYPMKPTIIPWRYYPQMDRTLLPSH-----TGTSGTPIINIHGTDPD 240
OY 190 -SPYCLE-YPPSQMLRTGNF-TFSYTFEEVPHSSYASQSILRLN--PLIDQYLYL 244
241 VQFYTIENSVPVHLRTGDEFATGTFEFDCKP--CRLTHWTQTRALGLPP-----FL 291
OY 245 NRTQNGSGAQNNDLFSRSPAGMSVQPKWLPGPCYRQGRSKTKTDNNNSNF----- 299
292 NSLPQSEGAATNFGDT-----GV-----QODKRGVTOGKNTNYITTEAT 329
OY 300 ---TWGASKYVNLNGRESIINP-----GTAMASHKDEDEKFPMSGVMLFGESAGAS 350
330 IMRPAEVGYAPYSPASIQGPFKPIAAGCGAQGDENQAADGNRVAFGHQGKTT 389
OY 351 TALD-----NMTIDEERIKATNPVATREFTGVAVNFQ 383
390 TTGETPERFTYIAHODTGRYPBGDWIONTINFNPVINDNVLPLTDPG---GKTGINY- 444
OY 384 SSSTDPATGVDVHMGALPGWVWQDRDVLQGPIMAKIPHTDGHFHSPLMGFGFLK--- 439
445 -TNIFNTYGPPLTALNNVP-----PYRNGQIMDEKFDTD-----LKRRLH 483
OY 440 -----NPPQILIKNTPVAN--PPAESATKFAFPITQYSTGVSEIEMELQK 487
484 VNAPFVCQNCPCQLPFKVAPNLTNQYDPPASANMSR---ITYSGFWMKGLVFXKL 539
OY 488 ENSKRANPEVQYTSNYSKASAVDFTVDNGLYTEPRPIG 526
DB 540 RASHTWNPIDQMSIN-----VDNQENYV-PSNIG 567
RESULT 15
COAT_FPV19 STANDARD; PRT; 727 AA.
ID COAT_FPV19
AC P24840;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2].
OS Feline panleukopenia virus (strain 193) (FPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_Taxid=10787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=193/70;
RX MEDLINE=91073139; PubMed=2174965;
RA Martyn J.C., Davidson B.E., Studeart M.J.;
RT "Nucleotide sequence of feline panleukopenia virus: comparison with
RL canine parvovirus identifies host-specific differences.";
RN J. Gen. Virol. 71:2747-2753 (1990).
RP [2]
RC SEQUENCE FROM N.A.
RP STRAIN=CU-4;

```

RX MEDLINE=91272479; PubMed=1647068;
RA Parish C.R.;
RT Mapping specific functions in the capsid structure of canine
RT parvovirus and feline panleukopenia virus using infectious plasmid
RT clones.
RL Virology 183:195-205(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55115; CA38911.1; -
DR EMBL; M38246; AAC37928.1; -
DR EMBL; M38246; AAC37929.1; -
DR PIR; B36608; VCPVPF.
DR HSSP; P30129; 4D6V.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 727 COAT PROTEIN VP1.
FT FT 144 727 COAT PROTEIN VP2.
FT FT 144 727 GLY-RICH.
FT FT 165 190 GLY-RICH.
FT FT 168 190 GLY-RICH.
FT FT 168 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 165 180 GLY-RICH.
SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621F55 CRC64;

Query Match 6.5%; Score 189; DB 1; Length 727;
Best Local Similarity 20.6%; Pred. No. 4.8e-06;
Matches 135; Conservative 73; Mismatches 216; Indels 232; Gaps 32;

QY 5 GGAPMDNNEGADGVGNAGNWHGDSITMLGDRVITITST-----RTWALPTVNN- 52
DB 153 GGQP-AVRNRRATGSGNGSGGGGGS--GSGVISTGTFFNNQTEFKFLNGWVEITANSS 209
QY 53 ---HL-----YKQI-----SSASTGASNDNHYFYSTPMGYDFDNRHCHSPRDW 95
DB 210 RLVLNMPESSENYKRVVNNMMDKTAVKGNMALDDIHQIVTWSLVDANAMGVWFNPGDW 269
QY 96 ORLLNNWGRPRKRLNKLFIQVKEYT---TNDGVTTIANLITVQVFSDEYOLPYV 152
DB 270 QLVNTMSELIHLVSEEOIFRVLKTVSESATQPTVYNNNDLTASLMLVLDSSNNTWFT 329
QY 153 LGSAGQCL-----PPFPA-----DVFMIPOGYLTLNNGSAVGRS----- 189
DB 330 PAMRSETLGFYPMKPIIPFWRKYFPQWDRILRSH-----TGTSTPTNVYHGTDPD 383
QY 190 -SFYCLE-YFPSQMLRTGNF--TF-----SYFEE-----VPHSYVHQSOL 229
DB 384 VQFYTIENSVPHLIRTGDEFATGTFFDCKPCRLTHTWQTRALGLPFLNLSIQSEGA 443
QY 230 DR-----LMNPLIDYL--YYLNRTONGS----- 251
DB 444 TNFEDIGVQDQRKRGVTQMGNTDYITATITMRPAEYVGSAPYSFEASTQGFKPTIAG 503
QY 252 -GSAQNKDLLFSRSGSPAGMSVQFQNLPGCY--RQORVSKTKTDNNNSNFTWTGASKY 307
DB 504 RGAQGTDBENQAAGDPR-----YAGRQHGQKTTTGTETPERFTY----- 543
QY 308 NLNGRESIINPGTAMASHKDEDEKFFPMGSGMITFGESAGASNTALD-NVMTIDEETIKA 366
DB 544 -----IAHODT-----GRAYEGDWIQINIFNLPTVNDVLLP 575

```

```

QY 367 TNVATERFGTVAVNFQSSSTDPATGVDVHMGALPGMWQDRDVYLQGPIMAKIPTHG 426
DB 576 TDPG-----GKGINV--INIFRTYGPLTALNNVP-----PYRNGQIMDEFPD 620
QY 427 FHPSPIMGGFGLK-----NPPQILIKTTPVPAN--PPAEFSATKFSFT 470
DB 621 -----LKPRLVNAPFVQCNNCPGQLFVKVAPMLTWEYDDPASANMSR----- 665
QY 471 QVSTGVSVIEWELQENSKRNPEVOYTSNVAKSANVDFYDNNGLYTERPPIG 526
DB 666 TYSDFWKKGLVKAKTLRASHHTNPIQWMSIN-----VDNQFNIV-PNNIG 710

```

Search completed: July 17, 2003, 18:32:07
Job time: 12.5714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 ; Search time 39.7143 Seconds
(without alignments)
2770.518 Million cell updates/sec

Title: US-09-807-802A-17
Perfect score: 2906
Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRVLRPL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	2906	100.0	736	12	Q9WBP8 adeno-ssoc
2	2884	99.2	736	12	Q9WBP8 adeno-ssoc
3	2530	87.1	736	12	Q56137 adeno-ssoc
4	2528	87.0	736	12	Q56139 adeno-ssoc
5	2481.5	85.4	533	12	Q65311 adeno-ssoc
6	2481.5	85.4	533	12	Q92917 adeno-ssoc
7	2481.5	85.4	533	12	Q56553 adeno-ssoc
8	1698	58.4	534	12	Q56552 adeno-ssoc
9	1698	58.4	534	12	Q67668 goose parvo
10	1698	58.4	534	12	Q67667 goose parvo
11	1697	58.4	732	12	Q67666 goose parvo
12	1690.5	58.2	732	12	Q8V395 goose parvo
13	1690	58.2	734	12	Q41855 adeno-ssoc
14	1676	57.7	534	12	Q83290 muscovy duc
15	1676	57.7	534	12	Q65446 barbare du
16	1676	57.7	732	12	Q65445 barbare du

17	1676	57.7	732	12	Q83289 muscovy duc
18	1665	57.3	724	12	Q9Y1J1 adeno-ssoc
19	1521	52.3	676	12	Q67672 goose parvo
20	715	24.6	179	12	Q9WA24 goose parvo
21	715	24.6	179	12	Q9WB07 duck parvo
22	711	24.5	179	12	Q9WN18 duck parvo
23	709	24.4	179	12	Q9WN19 duck parvo
24	707	24.3	179	12	Q9WN20 goose parvo
25	614	21.1	571	12	Q8QV4 minute vtru
26	614	21.1	703	12	Q8QV5 minute vtru
27	536.5	18.5	947	12	Q91809 bovine parv
28	501	17.2	554	12	Q9PZS9 human parvo
29	501	17.2	781	12	Q9PZT0 human parvo
30	495	17.0	785	12	Q9J0X4 pig-tailed
31	494.5	17.0	554	12	Q9J0P7 human parvo
32	494.5	17.0	781	12	Q9J0P8 human parvo
33	493	17.0	546	12	Q913X0 human parvo
34	493	17.0	773	12	Q913X1 human parvo
35	492.5	16.9	781	12	P89317 human parvo
36	492.5	16.9	781	12	P89318 human parvo
37	492	16.9	781	12	P89319 human parvo
38	490.5	16.9	781	12	P89316 human parvo
39	490	16.9	554	12	Q912B7 human eryth
40	490	16.9	554	12	Q90201 human parvo
41	490	16.9	769	12	Q9PZT4 human parvo
42	490	16.9	781	12	Q9JGS0 human eryth
43	490	16.9	781	12	Q912B8 human eryth
44	490	16.9	781	12	P90221 human parvo
45	490	16.9	781	12	P90222 human parvo

ALIGNMENTS

RESULT 1
Q9WBP8 PRELIMINARY; PRT; 736 AA.
ID Q9WBP8
AC Q9WBP8;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Capsid protein.
OS adeno-associated virus 1.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=85106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9214338; PubMed=10196295;
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;
RT "Gene therapy vectors based on adeno-associated virus type 1.";
RL J. Virol. 73:3994-4003 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Xiao W., Wilson J.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF063497; AAD27757.1; -;
DR InterPro: IPR001403; Parvo coat.
DR Pfam: PF00740; Parvo_coat.1.
SQ SEQUENCE 736 AA; 81375 MW; CFAFB9B85CD0595 CRC64;
Query Match 100.0%; Score 2906; DB 12; Length 736;
Best Local Similarity 100.0%; Pred. No. 7 4e-206;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASGGGAPMADNNEGADGVGNASGNHCDSTWIGDRVITTTSTRTMALPTYNHLYKQISS 60
DB 203 MASGGGAPMADNNEGADGVGNASGNHCDSTWIGDRVITTTSTRTMALPTYNHLYKQISS 262
QY 61 ASFGASNDNHYFGYSTPWGTFDENRFCHSPSPDQRLINNNGRFPKRLNFTLNIQVK 120
DB 263 ASFGASNDNHYFGYSTPWGTFDENRFCHSPSPDQRLINNNGRFPKRLNFTLNIQVK 322
QY 121 EVTTNDGVTTIANNLSTVQVFSDESEYQLPYVLSAHQGLPPFADVEMIPQYGLTLIN 180

```
Db 323 EVTTNDGTTTANNLSTVOVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQGYLTIN 382
Qy 181 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 240
Db 383 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 442
Qy 241 LYYLNRTQNGSGAONKDLLFSRGSPPAGMSVOPKMWLPGCYRQORVSKTKTDNNNSNF 300
Db 443 LYYLNRTQNGSGAONKDLLFSRGSPPAGMSVOPKMWLPGCYRQORVSKTKTDNNNSNF 502
Qy 301 WTGASKTNLNGRESIINPGTAMASHKODEDKFFPMGCVMIFGKESAGASNTALDNWMTD 360
Db 503 WTGASKTNLNGRESIINPGTAMASHKODEDKFFPMGCVMIFGKESAGASNTALDNWMTD 562
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHMGALPGVWODRDVYLOGPWAKI 420
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHMGALPGVWODRDVYLOGPWAKI 622
Qy 421 PHTDGHFHPSPPLMGSGFLKMPPOILLIKNTPVANPAEESATKFAFTIYQSTGOVSVE 480
Db 623 PHTDGHFHPSPPLMGSGFLKMPPOILLIKNTPVANPAEESATKFAFTIYQSTGOVSVE 682
Qy 481 IEMELOKENSKRNNPEVOYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 534
Db 683 IEMELOKENSKRNNPEVOYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 736
```

RESULT 2

```
AC 056137 PRELIMINARY; PRT; 736 AA.
ID 056137
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein VPI.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=68556;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2."
RT J. Virol. 72:309-319 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;
```

Query Match 99.2%; Score 2884; DB 12; Length 736;
Best Local Similarity 99.1%; Pred. No. 3.1e-204;
Matches 529; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTTMAALPTYNHLYKQISS 60
Db 203 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTTMAALPTYNHLYKQISS 262
Qy 61 ASTGASNDNHYFGYSTPWGTFDNRHCHSPRDWQRLINNMGFRPKLNFKLFNIQVK 120
Db 263 ASTGASNDNHYFGYSTPWGTFDNRHCHSPRDWQRLINNMGFRPKLNFKLFNIQVK 322
Qy 121 EVTTNDGTTTANNLSTVOVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQGYLTIN 180
Db 323 EVTTNDGTTTANNLSTVOVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQGYLTIN 382
Qy 181 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 240
```

```
Db 383 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 442
Qy 241 LYYLNRTQNGSGAONKDLLFSRGSPPAGMSVOPKMWLPGCYRQORVSKTKTDNNNSNF 300
Db 443 LYYLNRTQNGSGAONKDLLFSRGSPPAGMSVOPKMWLPGCYRQORVSKTKTDNNNSNF 502
Qy 301 WTGASKTNLNGRESIINPGTAMASHKODEDKFFPMGCVMIFGKESAGASNTALDNWMTD 360
Db 503 WTGASKTNLNGRESIINPGTAMASHKODEDKFFPMGCVMIFGKESAGASNTALDNWMTD 562
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHMGALPGVWODRDVYLOGPWAKI 420
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTDPAATGVHMGALPGVWODRDVYLOGPWAKI 622
Qy 421 PHTDGHFHPSPPLMGSGFLKMPPOILLIKNTPVANPAEESATKFAFTIYQSTGOVSVE 480
Db 623 PHTDGHFHPSPPLMGSGFLKMPPOILLIKNTPVANPAEESATKFAFTIYQSTGOVSVE 682
Qy 481 IEMELOKENSKRNNPEVOYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 534
Db 683 IEMELOKENSKRNNPEVOYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 736
```

RESULT 3

```
AC 056139 PRELIMINARY; PRT; 736 AA.
ID 056139
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein VPI.
OS adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2."
RT J. Virol. 72:309-319 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028705; AAB95452.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81906 MW; DD523331AD5F0D70F CRC64;
```

Query Match 87.1%; Score 2530; DB 12; Length 736;
Best Local Similarity 86.0%; Pred. No. 3.9e-178;
Matches 460; Conservative 27; Mismatches 46; Indels 2; Gaps 2;

```
Qy 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTTMAALPTYNHLYKQISS 60
Db 203 MASGGGAPMADNNEGADGVGNSSGNWCHDSTWLGDRVITTSRTTMAALPTYNHLYKQISS 262
Qy 61 ASTGASNDNHYFGYSTPWGTFDNRHCHSPRDWQRLINNMGFRPKLNFKLFNIQVK 120
Db 263 AS-GASNDNHYFGYSTPWGTFDNRHCHSPRDWQRLINNMGFRPKLNFKLFNIQVK 321
Qy 121 EVTTNDGTTTANNLSTVOVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQGYLTIN 180
Db 322 EVTTNDGTTTANNLSTVOVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQGYLTIN 381
Qy 181 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 240
Db 382 NGSQAVGRSSFFCYCLEYFPSSQMLRTGNNTFSTYEEVPHFSSVYAHSGSLDRLNPLIDQY 441
Qy 241 LYYLNRTQ-NGSGAONKDLLFSRGSPPAGMSVOPKMWLPGCYRQORVSKTKTDNNNSNF 299
Db 442 LYYLNRTQGTTSSTTQSGTRLLFSQAGPQSWSLQARWMLPGPCYRQORLSKTANDNNNSNF 501
```

```

QY 300 TWTGASKYKINNGRESIINPGTAMASHKODEDKFPFPMGSMVIFGESAGANTALDNVMT 359
D 502 PWTAAKSYKHLNGRDSLVPNGPAMASHKODEBEKFPFPMHGNLIFKEGEGTASNAELDNVMT 561
QY 360 DEEEIKATNPVATERFGTAVAVNFQSSSTDPAQDVHAMGALPGMWODRUVYLOGPIMAK 419
D 562 DEEEIRTTNPVATEQYGTAVANNLOQSSNTAPRTTIVNOGALPGMWODRUVYLOGPIMAK 621
QY 420 IPHTDGHFHPSPMLMGFGKLPKPPQILIKNTPVANPPAFSAATKFASTFOYSTQGVSV 479
D 622 IPHTDGHFHPSPMLMGFGKLPKPPQILIKNTPVANPPAFSAATKFASTFOYSTQGVSV 681
QY 480 EIEWELQKSKRNWPNVQYTSNVAKSANVDFVDDNNGLYTEPRPIGTRYLTP 534
D 682 EIEWELQKSKRNWPNVQYTSNVAKSANVDFVDDNNGLYTEPRPIGTRYLTP 736

```

RESULT 4

```

ID 065311 PRELIMINARY; PRT; 736 AA.
AC 065311;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 3.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=46350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RX MEDLINE=96266430; PubMed=8661429.
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
RT "Nucleotide sequencing and generation of an infectious clone of adeno-
RL associated virus 3."
RL Virology 221:208-217(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RA Muramatsu S., Brown K.E.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U48704; AAC55049.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat; 1.
SQ SEQUENCE 736 AA; 81660 MW; APLBPA7B5C67A10 CRC64;

```

Query Match 87.0%; Score 2528; DB 12; Length 736;
 Best Local Similarity 85.8%; Pred. No. 5.5e-178;
 Matches 459; Conservative 28; Mismatches 46; Indels 2; Gaps 2;

```

QY 1 MASGGGAPMADNNEGADGVGNASGNHCDSTWLGDRVITTSRTWALPTYNHLYKOISS 60
D 203 MASGGGAPMADNNEGADGVGNSSGNHCDSTWLGDRVITTSRTWALPTYNHLYKOISS 262
QY 61 ASAGASNDNHYFGYSTPMGWYFDENRFCHFSPPRDMQRLINNMGFRPKLNFPLNIQYK 120
D 263 QS-GASNDNHYFGYSTPMGWYFDENRFCHFSPPRDMQRLINNMGFRPKLNFPLNIQYK 321
QY 121 EYTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 180
D 322 GYVQNGGTTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 381
QY 181 NSGOAVGRSSFYCLEIFPSCMLRTGNFTFSYFEEVPPHSSYAHQSGLDRLANPLIDY 240
D 382 NSGOAVGRSSFYCLEIFPSCMLRTGNFTFSYFEEVPPHSSYAHQSGLDRLANPLIDY 441
QY 241 LYLINRTQ-NSGOAQNKDLIFSRGSPAGMSVOPKMWLPGPCYRQGRVSKTKTDNNNSNF 239
D 442 LYLINRTQ-NSGOAQNKDLIFSRGSPAGMSVOPKMWLPGPCYRQGRVSKTKTDNNNSNF 501
QY 300 TWTGASKYKINNGRESIINPGTAMASHKODEDKFPFPMGSMVIFGESAGANTALDNVMT 359

```

```

D 502 PWTAAKSYKHLNGRDSLVPNGPAMASHKODEBEKFPFPMHGNLIFKEGEGTASNAELDNVMT 561
QY 360 DEEEIKATNPVATERFGTAVAVNFQSSSTDPAQDVHAMGALPGMWODRUVYLOGPIMAK 419
D 562 DEEEIRTTNPVATEQYGTAVANNLOQSSNTAPRTTIVNOGALPGMWODRUVYLOGPIMAK 621
QY 420 IPHTDGHFHPSPMLMGFGKLPKPPQILIKNTPVANPPAFSAATKFASTFOYSTQGVSV 479
D 622 IPHTDGHFHPSPMLMGFGKLPKPPQILIKNTPVANPPAFSAATKFASTFOYSTQGVSV 681
QY 480 EIEWELQKSKRNWPNVQYTSNVAKSANVDFVDDNNGLYTEPRPIGTRYLTP 534
D 682 EIEWELQKSKRNWPNVQYTSNVAKSANVDFVDDNNGLYTEPRPIGTRYLTP 736

```

RESULT 5

```

ID 092917 PRELIMINARY; PRT; 533 AA.
AC 092917;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Major coat protein VP3.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; PubMed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif."
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschin J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043303; AAC03779.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein.
FT VARIANT 351 351 D -> N.
FT VARIANT 365 365 T -> N.
FT VARIANT 475 476 QV -> HV.
FT VARIANT 508 508 V -> R.
SQ SEQUENCE 533 AA; 60063 MW; 9E4DBC25810D4F0 CRC64;

```

Query Match 85.4%; Score 2481.5; DB 12; Length 533;
 Best Local Similarity 83.3%; Pred. No. 9.4e-175;
 Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

```

QY 1 MASGGGAPMADNNEGADGVGNASGNHCDSTWLGDRVITTSRTWALPTYNHLYKOISS 60
D 1 MASGGGAPMADNNEGADGVGNSSGNHCDSTWLGDRVITTSRTWALPTYNHLYKOISS 60
QY 61 ASAGASNDNHYFGYSTPMGWYFDENRFCHFSPPRDMQRLINNMGFRPKLNFPLNIQYK 120
D 61 QS-GASNDNHYFGYSTPMGWYFDENRFCHFSPPRDMQRLINNMGFRPKLNFPLNIQYK 119
QY 121 EYTTNDGVTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 180
D 120 EYVQNGGTTITANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVPMIPOYGYLTIN 179
QY 181 NSGOAVGRSSFYCLEIFPSCMLRTGNFTFSYFEEVPPHSSYAHQSGLDRLANPLIDY 240
D 180 NSGOAVGRSSFYCLEIFPSCMLRTGNFTFSYFEEVPPHSSYAHQSGLDRLANPLIDY 239
QY 241 LYLINRTQ-NSGOAQNKDLIFSRGSPAGMSVOPKMWLPGPCYRQGRVSKTKTDNNNSNF 300
D 240 LYLINRTQ-NSGOAQNKDLIFSRGSPAGMSVOPKMWLPGPCYRQGRVSKTKTDNNNSNF 299

```

Qy	301	WTGSAKYLNGREES.IINGGTAMASHKDDKEFFPMGSAVIFPKESAGASNTLADVMITD	360
Db	300	WTGATKTHLNGRDSL.VYNDEGPAMASHKDDKEFFPMGSAVIFPKESAKTNDIEKMTD	359
Qy	361	EEB1KATNPVATERPGTVAVNFQSSSTDPATGDVAMGALPGWVMQDRDVLQGIIMAKI	420
Db	360	EEB1RTTPVATEQYGSVSTNLDQRNGRQAATADVDTQGVLPWMQDRDVLQGIIMAKI	419
Qy	421	PHTOCHFPSPMLMGGEFGKKNPPOLLIKNTPPANPAPAEFSATKFASTFYQSTQSVSE	480
Db	420	PHTDHFPSPMLMGGEFGHLPOLLIKNTPPANPAPSTTFSAKPAASFYQSTQSVSE	479
Qy	481	IEMELQKNSKRMNPEVOYTSNYASANDVFVDMNGLYTEBRPIGTRETLREPL	534
Db	480	IEMELQKNSKRMNPEIQTSTNYNSVANDFVDMNGVSEBRPIGTRETLRENTL	533
RESULT 6			
ID	056653	PRELIMINARY;	PRT; 598 AA.
AC	056653;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Major coat protein VP2.		
OS	Adeno-associated virus 2 (AAV2).		
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.		
NC	NCBI_TaxId=10804;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95088582; PubMed=7996133;		
RA	Ruffing M., Heid H., Kleinschmidt J.A.;		
RT	"Mutations in the carboxy terminus of adeno-associated virus 2 capsid		
RT	proteins affect viral infectivity: lack of an RGD integrin-binding		
RT	motif."		
RL	J. Gen. Virol. 75:0-0(0).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,		
RA	Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,		
RA	Tratschkin J.-D., Weitz M.;		
RL	Submitted (JAN-1998) to the EMBL/GenBank/DDbJ databases.		
DR	EMBL; AF043303; AAC03778.1; --		
DR	InterPro; IPR01403; Parvo_coat.		
DR	Pfam; PF00740; Parvo_coat; 1.		
KW	Coat protein.		
FT	16		
FT	VARIANT 416 416 D -> N.		
FT	VARIANT 430 430 T -> N.		
FT	VARIANT 540 541 QV -> HV.		
FT	VARIANT 573 573 V -> R.		
SO	SEQUENCE 598 AA; 66619 MW; 070811ED9368B934 CRC64;		
Query Match			
Best Local Similarity 85.4%; Score 2481.5; DB 12; Length 598;			
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1			
Qy	1	MASGGAPMADNNEGADGVGNASGMWCDSTWLADRVITTSRTYALPTTNNHLTKYQIS	60
Db	66	MATGGAPMADNNEGADGVGNSSGMWCHDSTWNGRVITTSRTYALPTTNNHLTKYQIS	125
Qy	61	ASTGASNNHHFYGYSTPMGYDPENRPHCHGFSRDMQORLNNHNGRPKRLNFKLFNIOYK	120
Db	126	QS-GASNDNHIFGYSTPMGYDPENRPHCHGFSRDMQORLNNHNGRPKRLNFKLFNIOYK	184
Qy	121	EVTINDGVTTIANLNTSTVQVSDSEYQLPYVLGSAHQCLPPPADVFMIDQYGLTLN	180
Db	185	EVTQNDGVTTIANLNTSTVQVFTDSEYQLPYVLGSAHQCLPPPADVFMVQYGLTLN	244
Qy	181	NGSQAVGSSSPFCELEFPSPQMLRTGNPFSTTFEEVPHSHSYASQSIDRLANPLIDQY	240
Db	245	NGSQAVGSSSPFCELEFPSPQMLRTGNPFSTTFEEVPHSHSYASQSIDRLANPLIDQY	304

Qy		241	LYUNLRPNONGSGKXONKDILFSSRSPPAGMSVQPKWLLPGPCYRQARSKITDNNNSFT	300
Dd		305	LYIYSKRINP8GTTTQRSLQPSQAGADIRQSNNMLPGPCYRQARSKTSADNNSSEYS	364
Qy		301	WTGASKYNLNGRESIIINPGTAMASHKDEDEKFFPM5GMVFIFGESAGASNTALDNMTTD	360
Dd		365	WTGATKYTHLNGROSLAVNPGRPMASHKHDEEKFPPQSGLVIFGKGSEKTIVDIIEKWMTD	424
Qy		361	EELIKATNPVATERFQGVANVFOSSSTPDPAQGDVHAMCALRCGMWDODVDYLQGIIMKI	420
Dd		425	EEELRTNPVATEGYGSVTNLQGNRQAAPADVNTQVLGGMWODVDVYLQGIIMKI	484
Qy		421	PHTGHHPSPLMGFGFLKNPPQILLIKNTVPANMPAEPGFASFPIITQYSTQSVGE	480
Dd		485	PHIDGHHHPSPLMGFGFLKNPPQILLIKNTVPANMPTTFSAAFASFIITQYSTQSVGE	544
Qy		481	IEMELOKENSRRNPEVOYTSNVAKSANDVFTVNNGLYTEPRPIGTIRYLTPL	534
Dd		545	IEMELOKENSRRNPEIQYTSNVKSNVVDFVTDTNGVSSPPRGTYLTPL	598
<hr/>				
RESULT 7				
ID	OS6652	PRELIMINARY;	PRT;	735 AA.
AC	OS6652;			
DT	01-JUN-1998	(TREMBLrel. 06, Created)		
DT	01-JUN-1998	(TREMBLrel. 06, last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, last annotation update)		
DE	Major coat protein VPI.			
OC	Adeno-associated virus 2 (AAV2).			
CC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
OX	NCBI_TaxId=10804;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95088582; PubMed=7996133;			
RA	Ruffing M., Heid H., Kleinschmidt J.A.;			
RT	"Mutations in the carboxy terminus of adeno-associated virus 2 capsid			
RT	proteins affect viral infectivity: lack of an RGD integrin-binding			
RT	motif."			
RL	J. Gen. Virol. 75:0-0(0).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,			
RA	Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,			
RA	Tratschin J.-D., Weitz M.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF043303; AAC03780.1;			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
KW	Coat protein.			
FT	VARIA NT	76	D -> V.	
FT	VARIA NT	553	D -> N.	
FT	VARIA NT	567	T -> N.	
FT	VARIA NT	677	QV -> HV.	
FT	VARIA NT	710	V -> R.	
SO	SEQUENCE	735 AA;	81944 MM;	980BEEFA6908390B CRC64;
<hr/>				
Query Match 85.4%; Score 2481.5; DB 12; Length 735;				
Best Local Similarity 83.3%; Pred. No. 1.5e-174;				
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps				
Qy		1	MASGGCA PMADNNNGADGVGASGNMHCDSTWLGDVRVTTSTRTVALPYTNHLYKQISS	60
Dd		203	MATSSGAPMADNNNGAGVGSSGNMHCHDSITMMGDRIITTSTRITALPYTNHLYKQISS	262
Qy		61	ASTGASNDNHFGYSTFWGYFDENRFCHFSPRDMORLLINNMGFRPKLNFKLFNIQVK	120
Dd		263	QS-CASNDNHFGYSTFWGYFDENRFCHFSPRDMORLLINNMGFRPKLNFKLFNIQVK	321
Qy		121	EVTINDGVTTIANNLSTVQVFPSSSEVOLPVYLGSAHQGCIPPPADVFMIPOGYLT LN	180
Dd		322	EVTINDGVTTIANNLSTVQVFTTSEEQOLPVYLGSAHQGCIPPPADVFMIPOGYLT LN	381

```

QY 181 NSGAQVGRSSFYCLEYFPSCMLRTGNNFTSEYEEVPHSHSYAHQSOLDRLMPLIDQY 240
Db 382 NSGAQVGRSSFYCLEYFPSCMLRTGNNFTSEYEEVPHSHSYAHQSOLDRLMPLIDQY 441
QY 241 LYIANTONOGSAONKDLFSRGS PAMGMSVQPKMLPGPCYRQORV-SKTKTDN 300
Db 442 LYIANTONOGSAONKDLFSRGS PAMGMSVQPKMLPGPCYRQORV-SKTKTDN 501
QY 301 WTGASKNLNGRESIINPGTAMASHKDEKPEFPMGSMI FKGESASATLADNWTID 360
Db 502 WTGATKTHLNGRSLVNPGRPMASHKDEKPEFPMGSMI FKGESASATLADNWTID 561
QY 361 EEBIKATNPATERFGTVAVNFOSSTDPATGVDHAGALPGMWQORDVYLQGPWAKI 420
Db 562 EEBIKATNPATERFGTVAVNFOSSTDPATGVDHAGALPGMWQORDVYLQGPWAKI 621
QY 421 PHTDGHFHPSPMLMGFGKLPPOILIKNTVPANPPAESATYFASFTQYSTGVSE 480
Db 622 PHTDGHFHPSPMLMGFGKLPPOILIKNTVPANPPAESATYFASFTQYSTGVSE 681
QY 481 IEMELOKENSKRNPEVOYTSNYSKASANDVTUNNGLYTEPRIGRILTRPL 534
Db 682 IEMELOKENSKRNPEVOYTSNYSKASANDVTUNNGLYTEPRIGRILTRPL 735

RESULT 8
ID 067668 PRELIMINARY; PRT; 534 AA.
AC 067668;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VP3.
GN VP3.
OS goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RX MEDLINE=96010229; Pubmed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RA Zadori Z.;
RX EMBL; U25749; AAA83232.1; -.
DR InterPro; IPR001403; Parvo_coat.1.
DR Pfam; PF00740; Parvo_coat.1.
SQ SEQUENCE 534 AA; 5996 MW; F1F049558BEACE92 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 534;
Best Local Similarity 57.1%; Pred. No. 5.5e-117;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

QY 1 MASGGAGPAMADNNEGADGVGNASGNHCHDSITWLDGRVITTSRTWALPTYNNHLYQIIS 60
Db 1 MASGGAGPAMADNNEGADGVGNASGNHCHDSITWLDGRVITTSRTWALPTYNNHLYQIIS 60
QY 61 ASTGASNDNH--YFGYSTPMGYEDENRPHCHSPRDMQRLINNMWGRPRKLNFKLFINQ 118
Db 61 ASTGASNDNH--YFGYSTPMGYEDENRPHCHSPRDMQRLINNMWGRPRKLNFKLFINQ 117
QY 119 VKETVNDGVTITANNLTSTVOVFSDEYOLPYVLSAHOGLPFPADVFMIPOYGYLT 178
Db 119 VKETVNDGVTITANNLTSTVOVFSDEYOLPYVLSAHOGLPFPADVFMIPOYGYLT 177
QY 179 LN--NSGAQVGRSSFYCLEYFPSCMLRTGNNFTSEYEEVPHSHSYAHQSOLDRLMPL 235

```

```

Db 178 MHTNQGARNRDSARFYCLEYFPSCMLRTGNNFTSEYEEVPHSHSYAHQSOLDRLMPL 237
QY 236 LIDQYIYLNRTONOGSAONKDLFSRGS PAMGMSVQPKMLPGPCYRQORV-SKTKTDN 294
Db 238 LIDQYIYLNRTONOGSAONKDLFSRGS PAMGMSVQPKMLPGPCYRQORV-SKTKTDN 291
QY 295 NNSNFT-ATGASKNLNGRESIINPGTAMASHKDEKPEFPMGSMI FKGESASATLADNWTID 361
Db 292 NNSNFT-ATGASKNLNGRESIINPGTAMASHKDEKPEFPMGSMI FKGESASATLADNWTID 361
QY 352 ALDNNVITDEEBIKATNPATERFGTVAVNFOSSTDPATGVDHAGALPGMWQORDVYL 411
Db 352 ALDNNVITDEEBIKATNPATERFGTVAVNFOSSTDPATGVDHAGALPGMWQORDVYL 411
QY 412 LQGPWAKIPHTDGHFHPSPMLMGFGKLPPOILIKNTVPANPPAESATYFASFTQ 471
Db 412 LQGPWAKIPHTDGHFHPSPMLMGFGKLPPOILIKNTVPANPPAESATYFASFTQ 471
QY 472 YSTGOVSVLEIEMELOKENSKRNPEVOYTSNYSKASANDVTUNNGLYTEPRIGRILTRPL 531
Db 472 YSTGOVSVLEIEMELOKENSKRNPEVOYTSNYSKASANDVTUNNGLYTEPRIGRILTRPL 531
QY 532 RPL 534
Db 532 QNL 534

RESULT 9
ID 067667 PRELIMINARY; PRT; 587 AA.
AC 067667;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VP2.
GN VP2.
OS goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=38251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RX MEDLINE=96010229; Pubmed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RA Zadori Z.;
RX EMBL; U25749; AAA83231.1; -.
DR InterPro; IPR001403; Parvo_coat.1.
DR Pfam; PF00740; Parvo_coat.1.
SQ SEQUENCE 587 AA; 65246 MW; 0278E5D5FE7F0423 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 587;
Best Local Similarity 57.1%; Pred. No. 6.3e-117;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

QY 1 MASGGAGPAMADNNEGADGVGNASGNHCHDSITWLDGRVITTSRTWALPTYNNHLYQIIS 60
Db 1 MASGGAGPAMADNNEGADGVGNASGNHCHDSITWLDGRVITTSRTWALPTYNNHLYQIIS 60
QY 61 ASTGASNDNH--YFGYSTPMGYEDENRPHCHSPRDMQRLINNMWGRPRKLNFKLFINQ 118
Db 61 ASTGASNDNH--YFGYSTPMGYEDENRPHCHSPRDMQRLINNMWGRPRKLNFKLFINQ 113
QY 119 VKETVNDGVTITANNLTSTVOVFSDEYOLPYVLSAHOGLPFPADVFMIPOYGYLT 178
Db 119 VKETVNDGVTITANNLTSTVOVFSDEYOLPYVLSAHOGLPFPADVFMIPOYGYLT 170

```

Db 171 VKEVTTDQDTKIANNLSTIQVFTDDEHQLPYVLGSATGEMPPPSDVYALPOYGCT 230
Qy 179 LN---NSQAVGRSSFYCLEYFPSPOMLRTGNFTFSYFEEVPPHSSYAHQSGLDRMLNP 235
Db 231 MHTNONGARFNDRAFCLEYFPSPOMLRTGNFTFDFEEVPPHSMFAHSQDLDRMLNP 290
Qy 236 LIPOYLILNRTQNGSSAQNKDLFSRGSFAGMSVQPKMLPGPCYRQORV-SKTKTDN 294
Db 291 LVQGYLMNFEV-DSSRNAQ-----FKKAVGAGVGTGRNMLPGPKFLDQVRAVYGTGD 344
Qy 295 NNSNFT-WTGSCKNLNGRESIINPGTAMASHKODEDKFPPMSGVMIFGKE--SAGASNT 351
Db 345 NYANMNWMSNGNKNVNLKDRQYLLQPGVSAITYTBEGLSSLPADNIIIGIADPVRSSGTTA 404
Qy 352 ALDVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDY 411
Db 405 GISDIMEBQEVAPFTNGVMKPYGRVTNEQNTTAPTSDDLVLGALPGMWQNDIY 464
Qy 412 LQGPIMAKIPHTDGHFHPSPMLMGFGGLKNPPQILLIKNTVPANPPAEFSATKESPTIQ 471
Db 465 LQGPIGAKIPKTDGKHPSPMLGFGGLNPPQVFIKNTVPADPPVEYVHQKNSYITQ 524
Qy 472 YSTQGVSEIEMELOKENSKRANPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLT 531
Db 525 YSTQGVSEIEMELOKENSKRANPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLT 584
Qy 532 RPL 534
Db 585 QNL 587

RESULT 10

067666 PRELIMINARY; PRT; 732 AA.
ID 067666; AC 067666; DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VP1.
GN VP1.
OS goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=38251;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573 (1995).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRULENT B;
RA Zadori Z.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 732;
Best Local Similarity 57.1%; Pred. No. 8.6e-117;
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVNAGSNHCDSTWLGDRVITTSRTWALPTYNNHLYKOISS 60
Db 199 MAEGGGAGMGDSGSGADGVNAGSNHCDSTWLGDRVITTSRTWALPTYNNHLYKAITS 258
Qy 61 ASTGASNDN--HYGYSTPWGYDFNRFCHFSPRDWORLNNHNGRPRKLNKLNQ 118
Db 259 ---GTSQDANVQAGYSTPWGYDFNRFCHFSPRDWORLNNHNGRPRKLNKLNQ 315

Qy 119 VKEVTNDGVTIANNLSTVOYFSDSEYOLPYVLGSAGHQCPLPPFADYEMIPOXYLT 178
Db 316 VKEVTTDQDTKIANNLSTIQVFTDDEHQLPYVLGSATGEMPPPSDVYALPOYGCT 375
Qy 179 LN---NSQAVGRSSFYCLEYFPSPOMLRTGNFTFSYFEEVPPHSSYAHQSGLDRMLNP 235
Db 376 MHTNONGARFNDRAFCLEYFPSPOMLRTGNFTFDFEEVPPHSMFAHSQDLDRMLNP 435
Qy 236 LIPOYLILNRTQNGSSAQNKDLFSRGSFAGMSVQPKMLPGPCYRQORV-SKTKTDN 294
Db 436 LVQGYLMNFEV-DSSRNAQ-----FKKAVGAGVGTGRNMLPGPKFLDQVRAVYGTGD 489
Qy 295 NNSNFT-WTGSCKNLNGRESIINPGTAMASHKODEDKFPPMSGVMIFGKE--SAGASNT 351
Db 490 NYANMNWMSNGNKNVNLKDRQYLLQPGVSAITYTBEGLSSLPADNIIIGIADPVRSSGTTA 549
Qy 352 ALDVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDY 411
Db 550 GISDIMEBQEVAPFTNGVMKPYGRVTNEQNTTAPTSDDLVLGALPGMWQNDIY 609
Qy 412 LQGPIMAKIPHTDGHFHPSPMLMGFGGLKNPPQILLIKNTVPANPPAEFSATKESPTIQ 471
Db 610 LQGPIGAKIPKTDGKHPSPMLGFGGLNPPQVFIKNTVPADPPVEYVHQKNSYITQ 669
Qy 472 YSTQGVSEIEMELOKENSKRANPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLT 531
Db 670 YSTQGVSEIEMELOKENSKRANPEVOYTSNVAKSANVDFTVNNGLYTEPRPIGRYLT 729
Qy 532 RPL 534
Db 730 QNL 732

RESULT 11

08V395 PRELIMINARY; PRT; 732 AA.
ID 08V395; AC 08V395; DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Capsid protein VP.
GN VP.
OS goose parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
NCBI_TaxID=38251;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 732 AA; 81456 MW; 73F2E4BC769744B6 CRC64;

Query Match 58.4%; Score 1697; DB 12; Length 732;
Best Local Similarity 57.3%; Pred. No. 1e-116;
Matches 311; Conservative 82; Mismatches 132; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVNAGSNHCDSTWLGDRVITTSRTWALPTYNNHLYKOISS 60
Db 199 MAEGGGAGMGDSGSGADGVNAGSNHCDSTWLGDRVITTSRTWALPTYNNHLYKAITS 258
Qy 61 ASTGASNDN--HYGYSTPWGYDFNRFCHFSPRDWORLNNHNGRPRKLNKLNQ 118
Db 259 ---GTSQDANVQAGYSTPWGYDFNRFCHFSPRDWORLNNHNGRPRKLNKLNQ 315
Qy 119 VKEVTNDGVTIANNLSTVOYFSDSEYOLPYVLGSAGHQCPLPPFADYEMIPOXYLT 178
Db 316 VKEVTTDQDTKIANNLSTIQVFTDDEHQLPYVLGSATGEMPPPSDVYALPOYGCT 375

```

QY 179 LN---NGSOAVGRSSFYCLEFPSPOMLRTGNNFTSYTPEEVPFHSYAHOSLDRMLNP 235
DB 376 MHNONGARNDNSAFYCLEFPSPOMLRTGNNFTSYTPEEVPFHSYAHOSLDRMLNP 435
QY 236 LIDQYLNTNRONTONGSAGNOKDLFSGSPAGMSVOPKXNLPGCYRQORV-SKXTDN 294
DB 436 LVDQYLMNFNEV-DSRKAQ-----FKKAVKAGVGMGRMLDGPRLDQRYAVAYGTGD 489
QY 295 NNSNFT-WTGASKNYNGRESINPGTAMASHKODEKFPMSGVMTFGKE--SAGASNT 351
DB 490 NYANWNTWSNGNKNLDRQYLLQPGVPSATHTKEVASIIPAQVITGLADPYRSGSTTA 549
QY 352 ALNWMITDEBEIKATNPVATERPCTVAANVPOSSSTDPATGDVHAGALPGMWODDY 411
DB 550 GIDIMITDEBEIKATNPVATERPCTVAANVPOSSSTDPATGDVHAGALPGMWODDY 609
QY 412 LOGPIWAKIPHTDGHFHPSPFLMGFGGLKPNPQIILKNTVPANPAEFSATKFASTIQ 471
DB 610 LOGPIWAKIPHTDGHFHPSPFLMGFGGLKPNPQIILKNTVPANPAEFSATKFASTIQ 669
QY 472 YSTGQVSVEIEMELOKENSKRNWPEVQYTSNYAKSANDPTVDNNGLYTEPRPIGT 531
DB 670 YSSGQCTVEWVWELRKENSKRNWPEVQYTSNYAKSANDPTVDNNGLYTEPRPIGT 729
QY 532 RPL 534
DB 730 QNL 732

```

RESULT 12

```

ID 041855 PRELIMINARY; PRT; 734 AA.
AC 041855;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, last annotation update)
DE Capsid.
OS adeno-associated virus 4.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=57579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-646;
RX MEDLINE=97404695; PubMed=9261407;
RA "Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles.";
RT J. Virol. 71:6823-6833 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC VR-646;
RA "Cloning of AAV4, Yang L., Kotin R.M., Safer B.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89790; AAC58045.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat_1.
SQ SEQUENCE 734 AA; 80639 MW; 616CC27A777BB86F CRC64;

```

Query Match 58.2%; Score 1690.5; DB 12; Length 734;
 Best Local Similarity 59.4%; Pred. No. 3.1e-116;
 Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;

```

QY 2 ASGGGAPMADNNEGADGVGASGNMHCSTWLGDRVITTSRTWALPTTYNNHLYKQIS 61
DB 199 AAGGGAAY-BGGGAGADGVGASGNMHCSTWLGDRVITTSRTWALPTTYNNHLYKRLGE- 256
QY 62 STGASNDNHYFGYSPMGYFDPENRPHCFSPRDMQRLINNMGPRPKLNFPLFOYKE 121
DB 257 ---SLOSNTYNGSTPMGFDPENRPHCFSPRDMQRLINNMGPRPKLNFPLFOYKE 313
QY 122 VTNDGVTITANNLTSTVQVPSDSEYQLPYVLSAHQGLPFPADVEMIPQYGLTIN 178
DB 122 VTNDGVTITANNLTSTVQVPSDSEYQLPYVLSAHQGLPFPADVEMIPQYGLTIN 178

```

```

DB 314 VTSNGETTVANNLTSTVQVPSDSEYQLPYVMDAGEGSLPFPNDVFMVQYGCGLVT 373
QY 179 LNNGSOAVGRSSFYCLEFPSPOMLRTGNNFTSYTPEEVPFHSYAHOSLDRMLNP 235
DB 374 GNTSOQOTRNAPFYCLEFPSPOMLRTGNNFTSYTPEEVPFHSYAHOSLDRMLNP 433
QY 239 QYLYNNTNRONTONGSAGNOKDLFSGSPAGMSVOPKXNLPGCYRQORV-SKXTDN 294
DB 434 QYLMGLOSTTGTTLNAGNTTN---FTKLRPFNENFKKNLPGSIRKQGSKTA--N 488
QY 295 NNSNFTWGS---KY-----NLNGRESINPGTAMASHKODEKFPMSGVMTFGESAG 347
DB 489 QNYKIPATGSDSLIKETHTSTLDGRMSALTPGPMAATAGADSK-PSNSQLIAPGKONG 547
QY 348 ASNTALDNWITDEBEIKATNPVATERPCTVAANVPOSSSTDPATGDVHAGALPGMWOD 407
DB 548 NTATVPGLTILFTESEELAAATNATDTMGNLPGDQSNNSLPTVDELTAAGAIVGAWON 607
QY 408 RDVYLOGPIWAKIPHTDGHFHPSPFLMGFGGLKPNPQIILKNTVPANPAEFSATKFA 467
DB 608 RDVYLOGPIWAKIPHTDGHFHPSPFLMGFGGLKPNPQIILKNTVPANPAEFSATKFA 667
QY 468 FITQYSTGVSVIEEMELOKENSKRNWPEVQYTSNYAKSANDPTVDNNGLYTEPRPIGT 527
DB 668 FITQYSTGVSVQIDWEIOKERSKRNWPEVQYTSNYGQNSLLMADDAKYTEPRPIGT 727
QY 528 RYTRPL 534
DB 728 RYLTRPL 734

```

RESULT 13

```

ID 083290 PRELIMINARY; PRT; 732 AA.
AC 083290;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE Capsid protein.
OS Muscovy duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=37325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=89384/France;
RX MEDLINE=96406928; PubMed=8811015;
RA "Le Gall-Recul G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;
RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3) in a baculovirus expression system and demonstration of immunity induced by the recombinant protein.";
RT J. Gen. Virol. 77:2159-2163 (1996).
RL EMBL; Z68272; CA92575.1; -.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo coat_1.
FT CHAIN 146 732 VP2 CAPSID PROTEIN.
FT FT 199 732 VP3 CAPSID PROTEIN.
SQ SEQUENCE 732 AA; 81364 MW; DE70DCCAB215F4E2 CRC64;

```

Query Match 58.2%; Score 1690; DB 12; Length 732;
 Best Local Similarity 57.3%; Pred. No. 3.3e-116;
 Matches 310; Conservative 76; Mismatches 141; Indels 14; Gaps 6;

```

QY 1 MASGGAPMADNNEGADGVGASGNMHCSTWLGDRVITTSRTWALPTTYNNHLYKQIS 60
DB 199 MEGGSGAGADGVGASGNMHCSTWLGDRVITTSRTWALPTTYNNHLYKALTS 258
QY 61 ASGASNDNHYFGYSPMGYFDPENRPHCFSPRDMQRLINNMGPRPKLNFPLFOYKE 120
DB 259 GTNPDSN-TQYAGYSTPMGFDPENRPHCFSPRDMQRLINNMGPRPKLNFPLFOYKE 317
QY 121 EYVNDGVTITANNLTSTVQVPSDSEYQLPYVLSAHQGLPFPADVEMIPQYGLTIN 180
DB 318 EYVNDGVTITANNLTSTVQVPSDSEYQLPYVLSAHQGLPFPADVEMIPQYGLTIN 377

```

QY 181 ---NGSQAVERSSFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 237
DB 378 TNOSGARFNDRSAFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 437
QY 238 DQYLYLNRTONOGSAGNOKDLFPSRGSFAGMSVQPKMLPGPCYRQORVSKTK--TDNN 295
DB 438 DQYLYLNRTONOGSAGNOKDLFPSRGSFAGMSVQPKMLPGPCYRQORVSKTK--TDNN 295
QY 296 NSNFTWGAASKYNLNGRESIINPGTAMASHKODEDEKFFPMGSVMI FGEK--SAGASNTAL 353
DB 492 ANMSIWSKGNKVFLKDEREYLLQPGPVATHTTEDQASSVPANIIIGIADPYRSGSTIAGI 551
QY 354 DNWITDEEIKATNPATERFGTVAVNPOSSSTDPATGDVHAMGALPGWVODRDYLLQ 413
DB 552 SDIWDTEQGIAPNTNGVMPRYGLTVTNEQNTTAPNAELVYLGALPGWVODRDYLLQ 611
QY 414 GPIWAKIPTHGHHFSPMLGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITYOYS 473
DB 612 GPIWAKIPTHGHHFSPMLGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITYOYS 473
QY 474 TGOVSVEIEMELQENSKRNPNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGRYLLTRP 533
DB 672 TGOCTVEMWELRKENSKRNNPEIQTFSNFGNRTSTMFAPNETGGYVEDRLIGRITYLTON 731
QY 534 L 534
DB 732 L 732

RESULT 14

Q65446 PRELIMINARY; PRT; 534 AA.
AC 065446;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Capsid protein VP.
GN VP.
OS Barbardie duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
CX NCBI_TaxID=39118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FM;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573(1995).
DR EMBL; U22967; AAA83227.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat.1.
SQ SEQUENCE 534 AA; 59943 MW; 1963CED57BA70E0 CRC64;

Query Match 57.7%; Score 1676; DB 12; Length 534;

Best Local Similarity 56.9%; Pred. No. 2.3e-115; Indels 14; Gaps 6;
Matches 308; Conservative 77; Mismatches 142;

QY 1 MASGGAGPAMADNNAGADGVGNASGNWCHDSQMLGDRVITTSRTTWTALPTNNHLYKOISS 60
DB 1 MABGSGAMGDSAGAGDVGNASGNWCHDSQMLGDRVITTSRTTWTALPTNNHLYKOISS 60
QY 61 ASTGASNDNHFGYSTWPGYDFDNFRFCHFSPRDMQRLINNNGGFRPKLNFKLFNIQVK 120
DB 61 GTNPDSN--TOYAGSTWPGYDFDNFRFCHFSPRDMQRLINNNGGFRPKLNFKLFNIQVK 119
QY 121 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 180
DB 121 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 180
QY 120 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 179
DB 120 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 179
QY 181 ---NGSQAVERSSFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 237

DB 180 TNOSGARFNDRSAFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 239
QY 238 DQYLYLNRTONOGSAGNOKDLFPSRGSFAGMSVQPKMLPGPCYRQORVSKTK--TDNN 295
DB 240 DQYLYLNRTONOGSAGNOKDLFPSRGSFAGMSVQPKMLPGPCYRQORVSKTK--TDNN 295
QY 296 NSNFTWGAASKYNLNGRESIINPGTAMASHKODEDEKFFPMGSVMI FGEK--SAGASNTAL 353
DB 294 ANMSIWSKGNKVFLKDEREYLLQPGPVATHTTEDQASSVPANIIIGIADPYRSGSTIAGI 551
QY 354 DNWITDEEIKATNPATERFGTVAVNPOSSSTDPATGDVHAMGALPGWVODRDYLLQ 413
DB 354 SDIWDTEQGIAPNTNGVMPRYGLTVTNEQNTTAPNAELVYLGALPGWVODRDYLLQ 611
QY 414 GPIWAKIPTHGHHFSPMLGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITYOYS 473
DB 414 GPIWAKIPTHGHHFSPMLGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITYOYS 473
QY 474 TGOVSVEIEMELQENSKRNPNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGRYLLTRP 533
DB 474 TGOCTVEMWELRKENSKRNNPEIQTFSNFGNRTSTMFAPNETGGYVEDRLIGRITYLTON 731
QY 534 L 534
DB 534 L 534

RESULT 15

Q65445 PRELIMINARY; PRT; 587 AA.
AC 065445;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Capsid protein VP.
GN VP.
OS Barbardie duck parvovirus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
CX NCBI_TaxID=39118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FM;
RX MEDLINE=96010229; PubMed=7571426;
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;
RT "Analysis of the complete nucleotide sequences of goose and muscovy
RT duck parvoviruses indicates common ancestral origin with adeno-
RT associated virus 2.";
RL Virology 212:562-573(1995).
DR EMBL; U22967; AAA83226.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat.1.
SQ SEQUENCE 587 AA; 65198 MW; 2CE469DF3AC1086E CRC64;

Query Match 57.7%; Score 1676; DB 12; Length 587;

Best Local Similarity 56.9%; Pred. No. 2.6e-115; Indels 14; Gaps 6;
Matches 308; Conservative 77; Mismatches 142;

QY 1 MASGGAGPAMADNNAGADGVGNASGNWCHDSQMLGDRVITTSRTTWTALPTNNHLYKOISS 60
DB 54 MABGSGAMGDSAGAGDVGNASGNWCHDSQMLGDRVITTSRTTWTALPTNNHLYKOISS 113
QY 61 ASTGASNDNHFGYSTWPGYDFDNFRFCHFSPRDMQRLINNNGGFRPKLNFKLFNIQVK 120
DB 114 GTNPDSN--TOYAGSTWPGYDFDNFRFCHFSPRDMQRLINNNGGFRPKLNFKLFNIQVK 119
QY 121 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 180
DB 121 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 180
QY 173 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 232
DB 173 EVTTNDGVTIANNLTSTVQVFSDEYOLPYVLSAHQGLCPPADVFMIPOYGYLLTN 232
QY 181 ---NGSQAVERSSFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 237
DB 233 TNOSGARFNDRSAFYCLEYFPSSQMLRTGNNTFSYTEBEVPHSSYAHSSQSLDRLMPLI 292


```

Qy 238 DOYLTYLNRTONOSGAONKDLFFSRGSPAGMSVQPKWMLPGPCYRQORVSKTK--TDNN 235
Db 293 DOYLMNPFSEY-NGGRNQ-----FKKAVKGAFGAMGRNMLPGPKLDDORVRAISGTDNY 346
Qy 296 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKE--SAGASNTAL 353
Db 347 ANMSIMSKGNKVFLEKDBREYLLQPGPVATHTHEDQASSVPAONIIIGIAKDPYRSGSTLAGI 406
Qy 354 DNMTTDEEIKATNPVATERFGTVAVNFOSSSTDPATGDVHAMGALPGWVWQDRDYLO 413
Db 407 SDIMVTDEOEIAPLNGVGRPYGLTVTNEONTTAPTNABEVLVIGALPGWVWQDRDYLO 466
Qy 414 GPIWAKIPHTDGHFHPSPPLMGSGFGLKNPPOILIKNTPVANPAPAEFATKFASTFOYS 473
Db 467 GPIWAKIPKTDGKHPSPNIGFGLHNPPOVFIKNTFVPADPPLLEVNOKMSYITOYS 526
Qy 474 TGOVSVEIELELOKENSRRNPEVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRP 533
Db 527 TGOCTVEMVELREKNSKRWNPETIOFTSNFGNRTSTWFAENETGTYVEDRLIGTRYLTON 586
Qy 534 L 534
Db 587 L 587

```

Search completed: July 17, 2003, 18:34:33
 Job time : 41.7143 secs

